

# Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/US05/010913

International filing date: 31 March 2005 (31.03.2005)

Document type: Certified copy of priority document

Document details: Country/Office: US  
Number: 60/559,042  
Filing date: 01 April 2004 (01.04.2004)

Date of receipt at the International Bureau: 12 August 2005 (12.08.2005)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)



World Intellectual Property Organization (WIPO) - Geneva, Switzerland  
Organisation Mondiale de la Propriété Intellectuelle (OMPI) - Genève, Suisse

1352383

# THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

*August 02, 2005*

**THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE.**

**APPLICATION NUMBER: 60/559,042**

**FILING DATE: *April 01, 2004***

**RELATED PCT APPLICATION NUMBER: *PCT/US05/10913***



Certified by

Under Secretary of Commerce  
for Intellectual Property  
and Director of the United States  
Patent and Trademark Office

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

## PROVISIONAL APPLICATION FOR PATENT COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR 1.53(c).

**Express Mail Label No.** EL 961008104 US

INVENTOR(S)			
Given Name (first and middle [if any] )	Family Name or Surname	Residence (City and either State or Foreign Country)	
Steven	MAH	San Diego, California	
Andreas	BRAUN	San Diego, California	
Stefan M.	KAMMERER	San Diego, California	
Additional inventors are being named on the <u>1</u> separately numbered sheets attached hereto			
<b>TITLE OF THE INVENTION (500 characters max)</b>			
METHODS FOR IDENTIFYING RISK OF OSTEOARTHRITIS AND TREATMENTS THEREOF			
Direct all correspondence to: <b>CORRESPONDENCE ADDRESS</b>			
<input checked="" type="checkbox"/> Customer Number: <span style="border: 1px solid black; padding: 2px 20px;">25225</span>			
OR			
<input type="checkbox"/> Firm or Individual Name			
Address			
City		State	Zip
Country		Telephone	Fax
<b>ENCLOSED APPLICATION PARTS (check all that apply)</b>			
<input checked="" type="checkbox"/> Specification Number of Pages <span style="border: 1px solid black; padding: 2px 10px;">162</span> <input type="checkbox"/> CD(s), Number <span style="border: 1px solid black; padding: 2px 20px;"></span>			
<input checked="" type="checkbox"/> Drawing(s) Number of Sheets <span style="border: 1px solid black; padding: 2px 10px;">1</span> <input checked="" type="checkbox"/> Other <span style="border: 1px solid black; padding: 2px 50px;">Return Receipt Postcard</span>			
<input checked="" type="checkbox"/> Application Data Sheet. See 37 CFR 1.76 (4 pages) (specify):			
<b>METHOD OF PAYMENT OF FILING FEES FOR THIS PROVISIONAL APPLICATION FOR PATENT</b>			
<input checked="" type="checkbox"/> Applicant claims small entity status. See 37 CFR 1.27.			
<input type="checkbox"/> A check or money order is enclosed to cover the filing fees.			
<input checked="" type="checkbox"/> The Director is hereby authorized to charge filing fees or credit any overpayment to Deposit Account Number: <span style="border-bottom: 1px solid black; padding: 0 20px;">03-1952</span>			<b>FILING FEE AMOUNT (\$)</b>  <div style="border: 1px solid black; padding: 10px; width: 100px; margin: 0 auto;">80.00</div>
<input type="checkbox"/> Payment by credit card. Form PTO-2038 is attached.			
The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.			
<input checked="" type="checkbox"/> No <input type="checkbox"/> Yes, the name of the U.S. Government agency and the Government contract number are:			

*Respectfully submitted,*

[Page 1 of 2]

Date April 1, 2004

SIGNATURE \_\_\_\_\_  
TYPED OR  
PRINTED NAME

**Bruce D. Grant**

REGISTRATION NO. 47,608  
(if appropriate)

TELEPHONE

(858) 720-7962

Docket Number: 524593008900

**USE ONLY FOR FILING A PROVISIONAL APPLICATION FOR PATENT**

I hereby certify that this correspondence is being deposited with the U.S. Postal Service as Express Mail, Airbill No. EL 961008104 US, in an envelope addressed to: Mail Stop Provisional Patent Application, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450, on the date shown below.

Dated: 4/1/04 Signature:  (Deborah Wykes)

**PROVISIONAL APPLICATION COVER SHEET**  
**Additional Page**

PTO/SB/16 (08-03)

Approved for use through 07/31/06. OMB 0651-0032

U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

Docket Number 524593008900

INVENTOR(S)/APPLICANT(S)		
Given Name (first and middle [if any])	Family or Surname	Residence (City and either State or Foreign Country)
Matthew Roberts Rikard Henry Maria L.	NELSON RENELAND LANGDOWN	San Marcos, California San Diego, California San Diego, California

[Page 2 of 2]

## METHODS FOR IDENTIFYING RISK OF OSTEOARTHRITIS AND TREATMENTS THEREOF

### Field of the Invention

[0001] The invention relates to genetic methods for identifying risk of osteoarthritis and treatments that specifically target such diseases.

### Background

[0002] Osteoarthritis (OA) is a chronic disease usually affecting weight-bearing synovial joints. There are approximately 20 million Americans affected by OA and it is the leading cause of disability in the United States. In addition to extensive human suffering, OA also accounts for nearly all knee replacements and more than half of all hip replacements in the United States. Despite its prevalence, OA is poorly understood and there are few treatments available besides anti-inflammatory drugs and joint replacement.

[0003] Most commonly affecting middle-aged and older people, OA can range from very mild to very severe. It affects hands and weight-bearing joints such as knees, hips, feet and the back. Knee OA can be as disabling as any cardiovascular disease except stroke.

[0004] OA is characterized by the breakdown of cartilage in joints. Cartilage in joints cushions the ends of bones, and cartilage breakdown causes bones to rub against each other, causing pain and loss of movement. Type II collagen is the main component of cartilage, comprising 15-25% of the wet weight, approximately half the dry weight, and representing 90-95% of the total collagen content in the tissue. It forms fibrils that endow cartilage with tensile strength (Mayne, R. Arthritis Rheum. 32:241-246 (1989)).

### Summary

[0005] It has been discovered that certain polymorphic variations in human genomic DNA are associated with osteoarthritis. In particular, polymorphic variants in a locus containing a *ADAMTS2* region in human genomic DNA have been associated with risk of osteoarthritis.

[0006] Thus, featured herein are methods for identifying a subject at risk of osteoarthritis and/or a risk of osteoarthritis in a subject, which comprise detecting the presence or absence of one or more polymorphic variations associated with osteoarthritis in or around the loci described herein in a human nucleic acid sample. In an embodiment, two or more polymorphic variations are detected in two or more regions of which one is the *ADAMTS2* region. In certain embodiments, 3 or more, or 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20 or more polymorphic variants are detected.

[0007] Also featured are nucleic acids that include one or more polymorphic variations associated with occurrence of osteoarthritis, as well as polypeptides encoded by these nucleic acids. In addition, provided are methods for identifying candidate therapeutic molecules for treating osteoarthritis, as well as methods for treating osteoarthritis in a subject by identifying a subject at risk of osteoarthritis and treating the subject with a suitable prophylactic, treatment or therapeutic molecule.

[0008] Also provided are compositions comprising a cell from a subject having osteoarthritis or at risk of osteoarthritis and/or a *ADAMTS2* nucleic acid, with a RNAi, siRNA, antisense DNA or RNA, or ribozyme nucleic acid designed from a *ADAMTS2* nucleotide sequence. In an embodiment, the RNAi, siRNA, antisense DNA or RNA, or ribozyme nucleic acid is designed from a *ADAMTS2* nucleotide sequence that includes one or more polymorphic variations associated with osteoarthritis, and in some instances, specifically interacts with such a nucleotide sequence. Further, provided are arrays of nucleic acids bound to a solid surface, in which one or more nucleic acid molecules of the array have a *ADAMTS2* nucleotide sequence, or a fragment or substantially identical nucleic acid thereof, or a complementary nucleic acid of the foregoing. Featured also are compositions comprising a cell from a subject having osteoarthritis or at risk of osteoarthritis and/or a *ADAMTS2* polypeptide, with an antibody that specifically binds to the polypeptide. Thus, featured is an antibody that specifically binds to an epitope in the polypeptide that includes an amino acid encoded by a polymorphic site associated with osteoarthritis. In certain embodiments, the antibody specifically binds to an epitope comprising a valine or isoleucine encoded by rs398829 (e.g., an antibody that binds to an epitope comprising a valine at position 245 in an *ADAMTS2* polypeptide) A valine at position 245 is associated with increased risk of osteoarthritis.

#### Brief Description of the Drawings

[0009] Figure 1 shows proximal SNPs in a *ADAMTS2* region in genomic DNA. The position of each SNP in the chromosome is shown on the x-axis and the y-axis provides the negative logarithm of the p-value comparing the estimated allele to that of the control group. Also shown in the figure are exons and introns of the region in the approximate chromosomal positions.

#### Detailed Description

[0010] It has been discovered that a polymorphic variant in a locus containing a *ADAMTS2* region is associated with occurrence of osteoarthritis in subjects. Thus, detecting genetic determinants associated with an increased risk of osteoarthritis occurrence can lead to early identification of a predisposition to osteoarthritis and early prescription of preventative measures. Also, associating a *ADAMTS2* polymorphic variant with osteoarthritis has provided new targets for screening molecules useful in treatments of osteoarthritis.

### Osteoarthritis and Sample Selection

[0011] Osteoarthritis (OA), or degenerative joint disease, is one of the oldest and most common types of arthritis. It is characterized by the breakdown of the joint's cartilage. Cartilage is the part of the joint that cushions the ends of bones, and its breakdown causes bones to rub against each other, causing pain and loss of movement. Type II collagen is the main component of cartilage, comprising 15-25% of the wet weight, approximately half the dry weight, and representing 90-95% of the total collagen content in the tissue. It forms fibrils that endow cartilage with tensile strength (Mayne, R. Arthritis Rheum. 32:241-246 (1989)).

[0012] Most commonly affecting middle-aged and older people, OA can range from very mild to very severe. It affects hands and weight-bearing joints such as knees, hips, feet and the back. Knee OA can be as disabling as any cardiovascular disease except stroke. Whereas Ehlers-Danlos syndrome type VIIC is characterized by the retention of the N-terminal propeptide of type I collagen, osteoarthritis has been characterized by increased levels of type II collagen in osteoarthritic cartilage as measured by elevated C-propeptide concentrations (Nelson et al. (1998) J. Clin. Invest. 102(12):2115-2125).

[0013] Osteoarthritis affects an estimated 20.7 million Americans, mostly after age 45, with women more commonly affected than men. Physicians make a diagnosis of OA based on a physical exam and history of symptoms. X-rays are used to confirm diagnosis. Most people over 60 reflect the disease on X-ray, and about one-third have actual symptoms.

[0014] There are many factors that can cause OA. Obesity may lead to osteoarthritis of the knees. In addition, people with joint injuries due to sports, work-related activity or accidents may be at increased risk of developing OA.

[0015] Genetics has a role in the development of OA. Some people may be born with defective cartilage or with slight defects in the way that joints fit together. As a person ages, these defects may cause early cartilage breakdown in the joint or the inability to repair damaged or deteriorated cartilage in the joint.

[0016] Inclusion or exclusion of samples for an osteoarthritis pool may be based upon the following criteria: ethnicity (e.g., samples derived from an individual characterized as Caucasian); parental ethnicity (e.g., samples derived from an individual of British paternal and maternal descent); relevant phenotype information for the individual (e.g., case samples derived from individuals diagnosed with specific knee osteoarthritis (OA) and were recruited from an OA knee replacement clinic). Control samples may be selected based on relevant phenotype information for the individual (e.g., derived from individuals free of OA at several sites (knee, hand, hip etc)); and no family history of OA and/or rheumatoid arthritis. Additional phenotype information collected for both cases and controls may include age of the individual, gender, family history of OA, diagnosis with osteoarthritis (joint location of OA, date of primary diagnosis, age of individual as of primary diagnosis), knee history (current symptoms,

any major knee injury, meniscectomy, knee replacement surgery, age of surgery), HRT history, osteoporosis diagnosis.

[0017] Based in part upon selection criteria set forth above, individuals having osteoarthritis can be selected for genetic studies. Also, individuals having no history of osteoarthritis often are selected for genetic studies, as described hereafter.

#### Polymorphic Variants Associated with Osteoarthritis

[0018] A genetic analysis provided herein linked osteoarthritis with polymorphic variant nucleic acid sequences in the human genome. As used herein, the term “polymorphic site” refers to a region in a nucleic acid at which two or more alternative nucleotide sequences are observed in a significant number of nucleic acid samples from a population of individuals. A polymorphic site may be a nucleotide sequence of two or more nucleotides, an inserted nucleotide or nucleotide sequence, a deleted nucleotide or nucleotide sequence, or a microsatellite, for example. A polymorphic site that is two or more nucleotides in length may be 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 or more, 20 or more, 30 or more, 50 or more, 75 or more, 100 or more, 500 or more, or about 1000 nucleotides in length, where all or some of the nucleotide sequences differ within the region. A polymorphic site is often one nucleotide in length, which is referred to herein as a “single nucleotide polymorphism” or a “SNP.”

[0019] Where there are two, three, or four alternative nucleotide sequences at a polymorphic site, each nucleotide sequence is referred to as a “polymorphic variant” or “nucleic acid variant.” Where two polymorphic variants exist, for example, the polymorphic variant represented in a minority of samples from a population is sometimes referred to as a “minor allele” and the polymorphic variant that is more prevalently represented is sometimes referred to as a “major allele.” Many organisms possess a copy of each chromosome (*e.g.*, humans), and those individuals who possess two major alleles or two minor alleles are often referred to as being “homozygous” with respect to the polymorphism, and those individuals who possess one major allele and one minor allele are normally referred to as being “heterozygous” with respect to the polymorphism. Individuals who are homozygous with respect to one allele are sometimes predisposed to a different phenotype as compared to individuals who are heterozygous or homozygous with respect to another allele.

[0020] In genetic analysis that associate polymorphic variants with osteoarthritis, samples from individuals having osteoarthritis and individuals not having osteoarthritis often are allelotyped and/or genotyped. The term “allelotype” as used herein refers to a process for determining the allele frequency for a polymorphic variant in pooled DNA samples from cases and controls. By pooling DNA from each group, an allele frequency for each SNP in each group is calculated. These allele frequencies are then compared to one another. The term “genotyped” as used herein refers to a process for determining a



genotype of one or more individuals, where a “genotype” is a representation of one or more polymorphic variants in a population.

[0021] A genotype or polymorphic variant may be expressed in terms of a “haplotype,” which as used herein refers to two or more polymorphic variants occurring within genomic DNA in a group of individuals within a population. For example, two SNPs may exist within a gene where each SNP position includes a cytosine variation and an adenine variation. Certain individuals in a population may carry one allele (heterozygous) or two alleles (homozygous) having the gene with a cytosine at each SNP position. As the two cytosines corresponding to each SNP in the gene travel together on one or both alleles in these individuals, the individuals can be characterized as having a cytosine/cytosine haplotype with respect to the two SNPs in the gene.

[0022] As used herein, the term “phenotype” refers to a trait which can be compared between individuals, such as presence or absence of a condition, a visually observable difference in appearance between individuals, metabolic variations, physiological variations, variations in the function of biological molecules, and the like. An example of a phenotype is occurrence of osteoarthritis.

[0023] Researchers sometimes report a polymorphic variant in a database without determining whether the variant is represented in a significant fraction of a population. Because a subset of these reported polymorphic variants are not represented in a statistically significant portion of the population, some of them are sequencing errors and/or not biologically relevant. Thus, it is often not known whether a reported polymorphic variant is statistically significant or biologically relevant until the presence of the variant is detected in a population of individuals and the frequency of the variant is determined. Methods for detecting a polymorphic variant in a population are described herein, specifically in Example 2. A polymorphic variant is statistically significant and often biologically relevant if it is represented in 5% or more of a population, sometimes 10% or more, 15% or more, or 20% or more of a population, and often 25% or more, 30% or more, 35% or more, 40% or more, 45% or more, or 50% or more of a population.

[0024] A polymorphic variant may be detected on either or both strands of a double-stranded nucleic acid. Also, a polymorphic variant may be located within an intron or exon of a gene or within a portion of a regulatory region such as a promoter, a 5′ untranslated region (UTR), a 3′ UTR, and in DNA (*e.g.*, genomic DNA (gDNA) and complementary DNA (cDNA)), RNA (*e.g.*, mRNA, tRNA, and rRNA), or a polypeptide. Polymorphic variations may or may not result in detectable differences in gene expression, polypeptide structure, or polypeptide function.

[0025] It was determined that polymorphic variations associated with an increased risk of osteoarthritis existed in a *ADAMTS2* region in SEQ ID NO: 1. In certain embodiments, a polymorphic variant at position rs398829 in the human genome was associated with an increased risk of osteoarthritis, and in a specific embodiment, a guanine at position rs398829 was associated with an increased risk of osteoarthritis.

[0026] Polymorphic variants in and around the *ADAMTS2* locus were tested for association with osteoarthritis. These include polymorphic variants at positions in SEQ ID NO: 1 selected from the group consisting of 210, 3608, 3609, 4318, 5593, 5629, 5639, 5640, 8943, 17968, 19887, 21034, 21085, , 21596, 23379, 23432, 24007, 26121, 26273, 26755, 27411, 27710, 27842, 28379, 29603, 31232, 31504, 32583, 32794, 32840, 33044, 33150, 33218, 33513, 33959, 34486, 36289, 36570, 38247, 38477, 38518, 38529, 38667, 39781, 39856, 39927, 40506, 41869, 42452, 44788, 46059, 46846, 47712, 48796, 49441, 49602, 49723, 50050, 50171, 50477, 50818, 50833, 50881, 50882, 51386, 51534, 52317, 52368, 52970, 53023, 53356, 53882, 54553, 55475, 55530, 55691, 55848, 55879, 56316, 56911, 57320, 57391, 57437, 57478, 57500, 59111, 59333, 59715, 59804, 59851, 59929, 60052, 60240, 60359, 60381, 60456, 60724, 60875, 60968, 60978, 60998, 61557, 62091, 62645, 62943, 63131, 63145, 63406, 63427, 63554, 63661, 64093, 64153, 64409, 64544, 65257, 65626, 65739, 66392, 66720, 69177, 69336, 69636, 69823, 69928, 70547, 70633, 71805, 72181, 72200, 72474, 72567, 72973, 73468, 73889, 75730, 75970, 76114, 76342, 76449, 76465, 76791, 78042, 80758, 80778, 81356, 81576, 81689, 81759, 81950, 82562, 83591, 83700, 83821, 83842, 83923, 83929, 84021, 84175, 84417, 84747, 85746, 86129, 86335, 87315, 87648, 87764, 87770, 88221, 90474, 91148, 91150, 91160, 91733, 91772, 91785, 93140, 93148, 96080, 96157, 96313, 96759, 97026, 97320, 97732, 98713, 99707, 99959, 100009, 100020, 100065, 100086, 101270, 101276, 101371, 101376, 101439, 101820, 102392, 102602, 102604, 102896, 189104, 189134 and 189205. Polymorphic variants at the following positions in SEQ ID NO: 1 in particular were associated with an increased risk of osteoarthritis: 5640, 33150, 38247, 38529, 46846, 49723, 50050, 63427, 73889, 189104 and rs428901, where specific embodiments are directed to positions 46846, 73889, 189104 and/or rs428901. In particular, the following polymorphic variants in SEQ ID NO: 1 were associated with risk of osteoarthritis: a cytosine at position 5640, a cytosine at position 33150, an adenine at position 38247, a thymine at position 38529, an adenine at position 46846, a cytosine at position 49723, a cytosine at position 50050, a cytosine a position 63427, a guanine at position 73889, a thymine at position 189104, and an adenine at position rs428901.

[0027] Based in part upon analyses summarized in Figure 1, a region with significant association has been identified in a locus associated with osteoarthritis. Any polymorphic variants associated with osteoarthritis in a region of significant association can be utilized for embodiments described herein. For example, polymorphic variants in a region spanning chromosome positions 178746000 to 178751000 (approximately 5,000 nucleotides in length) in a *ADAMTS2* locus have significant association (chromosome positions are within NCBI's Genome build 34).

#### Additional Polymorphic Variants Associated with Osteoarthritis

[0028] Also provided is a method for identifying polymorphic variants proximal to an incident, founder polymorphic variant associated with osteoarthritis. Thus, featured herein are methods for

identifying a polymorphic variation associated with osteoarthritis that is proximal to an incident polymorphic variation associated with osteoarthritis, which comprises identifying a polymorphic variant proximal to the incident polymorphic variant associated with osteoarthritis, where the incident polymorphic variant is in a *ADAMTS2* nucleotide sequence. The nucleotide sequence often comprises a polynucleotide sequence selected from the group consisting of (a) a polynucleotide sequence of SEQ ID NO: 1-3; (b) a polynucleotide sequence that encodes a polypeptide having an amino acid sequence encoded by a polynucleotide sequence of SEQ ID NO: 1-3; and (c) a polynucleotide sequence that encodes a polypeptide having an amino acid sequence that is 90% or more identical to an amino acid sequence encoded by a nucleotide sequence of SEQ ID NO: 1-3 or a polynucleotide sequence 90% or more identical to the polynucleotide sequence of SEQ ID NO: 1-3. The presence or absence of an association of the proximal polymorphic variant with osteoarthritis then is determined using a known association method, such as a method described in the Examples hereafter. In an embodiment, the incident polymorphic variant is a polymorphic variant associated with osteoarthritis described herein. In another embodiment, the proximal polymorphic variant identified sometimes is a publicly disclosed polymorphic variant, which for example, sometimes is published in a publicly available database. In other embodiments, the polymorphic variant identified is not publicly disclosed and is discovered using a known method, including, but not limited to, sequencing a region surrounding the incident polymorphic variant in a group of nucleic samples. Thus, multiple polymorphic variants proximal to an incident polymorphic variant are associated with osteoarthritis using this method.

[0029] The proximal polymorphic variant often is identified in a region surrounding the incident polymorphic variant. In certain embodiments, this surrounding region is about 50 kb flanking the first polymorphic variant (*e.g.* about 50 kb 5' of the first polymorphic variant and about 50 kb 3' of the first polymorphic variant), and the region sometimes is composed of shorter flanking sequences, such as flanking sequences of about 40 kb, about 30 kb, about 25 kb, about 20 kb, about 15 kb, about 10 kb, about 7 kb, about 5 kb, or about 2 kb 5' and 3' of the incident polymorphic variant. In other embodiments, the region is composed of longer flanking sequences, such as flanking sequences of about 55 kb, about 60 kb, about 65 kb, about 70 kb, about 75 kb, about 80 kb, about 85 kb, about 90 kb, about 95 kb, or about 100 kb 5' and 3' of the incident polymorphic variant.

[0030] In certain embodiments, polymorphic variants associated with osteoarthritis are identified iteratively. For example, a first proximal polymorphic variant is associated with osteoarthritis using the methods described above and then another polymorphic variant proximal to the first proximal polymorphic variant is identified (*e.g.*, publicly disclosed or discovered) and the presence or absence of an association of one or more other polymorphic variants proximal to the first proximal polymorphic variant with osteoarthritis is determined.

[0031] The methods described herein are useful for identifying or discovering additional polymorphic variants that may be used to further characterize a gene, region or loci associated with a condition, a disease (e.g., osteoarthritis), or a disorder. For example, allelotyping or genotyping data from the additional polymorphic variants may be used to identify a functional mutation or a region of linkage disequilibrium. In certain embodiments, polymorphic variants identified or discovered within a region comprising the first polymorphic variant associated with osteoarthritis are genotyped using the genetic methods and sample selection techniques described herein, and it can be determined whether those polymorphic variants are in linkage disequilibrium with the first polymorphic variant. The size of the region in linkage disequilibrium with the first polymorphic variant also can be assessed using these genotyping methods. Thus, provided herein are methods for determining whether a polymorphic variant is in linkage disequilibrium with a first polymorphic variant associated with osteoarthritis, and such information can be used in prognosis/diagnosis methods described herein.

#### Isolated Nucleic Acids

[0032] Featured herein are isolated *ADAMTS2* nucleic acid variants depicted in SEQ ID NO: 1-3, and substantially identical nucleic acids thereof. A nucleic acid variant may be represented on one or both strands in a double-stranded nucleic acid or on one chromosomal complement (heterozygous) or both chromosomal complements (homozygous). *ADAMTS2* exists in two forms, a "long" form comprising a molecule approximately 130 kDa in length (e.g., SEQ ID NO: 2 for cDNA sequence and SEQ ID NO: 4 for amino acid sequence), and a "short" form comprising a molecule approximately 70 kDa in length (e.g., SEQ ID NO: 3 for cDNA sequence and SEQ ID NO: 5 for amino acid sequence). Provided herein are polynucleotide sequences encoding both the short and long forms of *ADAMTS2*.

[0033] As used herein, the term "nucleic acid" includes DNA molecules (e.g., a complementary DNA (cDNA) and genomic DNA (gDNA)) and RNA molecules (e.g., mRNA, rRNA, siRNA and tRNA) and analogs of DNA or RNA, for example, by use of nucleotide analogs. The nucleic acid molecule can be single-stranded and it is often double-stranded. The term "isolated or purified nucleic acid" refers to nucleic acids that are separated from other nucleic acids present in the natural source of the nucleic acid. For example, with regard to genomic DNA, the term "isolated" includes nucleic acids which are separated from the chromosome with which the genomic DNA is naturally associated. An "isolated" nucleic acid is often free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and/or 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated nucleic acid molecule can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of 5' and/or 3' nucleotide sequences which flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular

material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized. As used herein, the term “gene” refers to a nucleotide sequence that encodes a polypeptide.

[0034] Also included herein are nucleic acid fragments. These fragments often have a nucleotide sequence identical to a nucleotide sequence of SEQ ID NO: 1-3, a nucleotide sequence substantially identical to a nucleotide sequence of SEQ ID NO: 1-3, or a nucleotide sequence that is complementary to the foregoing. The nucleic acid fragment may be identical, substantially identical or homologous to a nucleotide sequence in an exon or an intron in a nucleotide sequence of SEQ ID NO: 1-3, and may encode a domain or part of a domain of a polypeptide. Sometimes, the fragment will comprises one or more of the polymorphic variations described herein as being associated with osteoarthritis. The nucleic acid fragment is often 50, 100, or 200 or fewer base pairs in length, and is sometimes about 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 2000, 3000, 4000, 5000, 10000, 15000, or 20000 base pairs in length. A nucleic acid fragment that is complementary to a nucleotide sequence identical or substantially identical to a nucleotide sequence in SEQ ID NO: 1-3 and hybridizes to such a nucleotide sequence under stringent conditions is often referred to as a “probe.” Nucleic acid fragments often include one or more polymorphic sites, or sometimes have an end that is adjacent to a polymorphic site as described hereafter.

[0035] An example of a nucleic acid fragment is an oligonucleotide. As used herein, the term “oligonucleotide” refers to a nucleic acid comprising about 8 to about 50 covalently linked nucleotides, often comprising from about 8 to about 35 nucleotides, and more often from about 10 to about 25 nucleotides. The backbone and nucleotides within an oligonucleotide may be the same as those of naturally occurring nucleic acids, or analogs or derivatives of naturally occurring nucleic acids, provided that oligonucleotides having such analogs or derivatives retain the ability to hybridize specifically to a nucleic acid comprising a targeted polymorphism. Oligonucleotides described herein may be used as hybridization probes or as components of prognostic or diagnostic assays, for example, as described herein.

[0036] Oligonucleotides are typically synthesized using standard methods and equipment, such as the ABI™3900 High Throughput DNA Synthesizer and the EXPEDITE™ 8909 Nucleic Acid Synthesizer, both of which are available from Applied Biosystems (Foster City, CA). Analogs and derivatives are exemplified in U.S. Pat. Nos. 4,469,863; 5,536,821; 5,541,306; 5,637,683; 5,637,684; 5,700,922; 5,717,083; 5,719,262; 5,739,308; 5,773,601; 5,886,165; 5,929,226; 5,977,296; 6,140,482; WO 00/56746; WO 01/14398, and related publications. Methods for synthesizing oligonucleotides comprising such analogs or derivatives are disclosed, for example, in the patent publications cited above

and in U.S. Pat. Nos. 5,614,622; 5,739,314; 5,955,599; 5,962,674; 6,117,992; in WO 00/75372; and in related publications.

[0037] Oligonucleotides may also be linked to a second moiety. The second moiety may be an additional nucleotide sequence such as a tail sequence (*e.g.*, a polyadenosine tail), an adapter sequence (*e.g.*, phage M13 universal tail sequence), and others. Alternatively, the second moiety may be a non-nucleotide moiety such as a moiety which facilitates linkage to a solid support or a label to facilitate detection of the oligonucleotide. Such labels include, without limitation, a radioactive label, a fluorescent label, a chemiluminescent label, a paramagnetic label, and the like. The second moiety may be attached to any position of the oligonucleotide, provided the oligonucleotide can hybridize to the nucleic acid comprising the polymorphism.

#### Uses for Nucleic Acid Sequence

[0038] Nucleic acid coding sequences may be used for diagnostic purposes for detection and control of polypeptide expression. Also, included herein are oligonucleotide sequences such as antisense RNA, small-interfering RNA (siRNA) and DNA molecules and ribozymes that function to inhibit translation of a polypeptide. Antisense techniques and RNA interference techniques are known in the art and are described herein.

[0039] Ribozymes are enzymatic RNA molecules capable of catalyzing the specific cleavage of RNA. The mechanism of ribozyme action involves sequence specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, hammerhead motif ribozyme molecules may be engineered that specifically and efficiently catalyze endonucleolytic cleavage of RNA sequences corresponding to or complementary to *ADAMTS2* nucleotide sequences. Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites which include the following sequences, GUA, GUU and GUC. Once identified, short RNA sequences of between fifteen (15) and twenty (20) ribonucleotides corresponding to the region of the target gene containing the cleavage site may be evaluated for predicted structural features such as secondary structure that may render the oligonucleotide sequence unsuitable. The suitability of candidate targets may also be evaluated by testing their accessibility to hybridization with complementary oligonucleotides, using ribonuclease protection assays.

[0040] Antisense RNA and DNA molecules, siRNA and ribozymes may be prepared by any method known in the art for the synthesis of RNA molecules. These include techniques for chemically synthesizing oligodeoxyribonucleotides well known in the art such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA sequences encoding the antisense RNA molecule. Such DNA sequences may be incorporated

into a wide variety of vectors which incorporate suitable RNA polymerase promoters such as the T7 or SP6 polymerase promoters. Alternatively, antisense cDNA constructs that synthesize antisense RNA constitutively or inducibly, depending on the promoter used, can be introduced stably into cell lines.

[0041] DNA encoding a polypeptide also may have a number of uses for the diagnosis of diseases, including osteoarthritis, resulting from aberrant expression of a target gene described herein. For example, the nucleic acid sequence may be used in hybridization assays of biopsies or autopsies to diagnose abnormalities of expression or function (*e.g.*, Southern or Northern blot analysis, *in situ* hybridization assays).

[0042] In addition, the expression of a polypeptide during embryonic development may also be determined using nucleic acid encoding the polypeptide. As addressed, *infra*, production of functionally impaired polypeptide is the cause of various disease states, such as osteoarthritis. *In situ* hybridizations using polypeptide as a probe may be employed to predict problems related to osteoarthritis. Further, as indicated, *infra*, administration of human active polypeptide, recombinantly produced as described herein, may be used to treat disease states related to functionally impaired polypeptide. Alternatively, gene therapy approaches may be employed to remedy deficiencies of functional polypeptide or to replace or compete with dysfunctional polypeptide.

#### Expression Vectors, Host Cells, and Genetically Engineered Cells

[0043] Provided herein are nucleic acid vectors, often expression vectors, which contain a *ADAMTS2* nucleotide sequence, or a substantially identical sequence thereof. As used herein, the term “vector” refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked and can include a plasmid, cosmid, or viral vector. The vector can be capable of autonomous replication or it can integrate into a host DNA. Viral vectors may include replication defective retroviruses, adenoviruses and adeno-associated viruses for example.

[0044] A vector can include a *ADAMTS2* nucleotide sequence in a form suitable for expression of an encoded target polypeptide or target nucleic acid in a host cell. A “target polypeptide” is a polypeptide encoded by a *ADAMTS2* nucleotide sequence, or a substantially identical nucleotide sequence thereof. The recombinant expression vector typically includes one or more regulatory sequences operatively linked to the nucleic acid sequence to be expressed. The term “regulatory sequence” includes promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence, as well as tissue-specific regulatory and/or inducible sequences. The design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of polypeptide desired, and the like. Expression vectors can be introduced into host cells to produce target polypeptides, including fusion polypeptides.

[0045] Recombinant expression vectors can be designed for expression of target polypeptides in prokaryotic or eukaryotic cells. For example, target polypeptides can be expressed in *E. coli*, insect cells (e.g., using baculovirus expression vectors), yeast cells, or mammalian cells. Suitable host cells are discussed further in Goeddel, *Gene Expression Technology: Methods in Enzymology 185*, Academic Press, San Diego, CA (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

[0046] Expression of polypeptides in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion polypeptides. Fusion vectors add a number of amino acids to a polypeptide encoded therein, usually to the amino terminus of the recombinant polypeptide. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant polypeptide; 2) to increase the solubility of the recombinant polypeptide; and 3) to aid in the purification of the recombinant polypeptide by acting as a ligand in affinity purification. Often, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant polypeptide to enable separation of the recombinant polypeptide from the fusion moiety subsequent to purification of the fusion polypeptide. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith & Johnson, *Gene 67*: 31-40 (1988)), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding polypeptide, or polypeptide A, respectively, to the target recombinant polypeptide.

[0047] Purified fusion polypeptides can be used in screening assays and to generate antibodies specific for target polypeptides. In a therapeutic embodiment, fusion polypeptide expressed in a retroviral expression vector is used to infect bone marrow cells that are subsequently transplanted into irradiated recipients. The pathology of the subject recipient is then examined after sufficient time has passed (e.g., six (6) weeks).

[0048] Expressing the polypeptide in host bacteria with an impaired capacity to proteolytically cleave the recombinant polypeptide is often used to maximize recombinant polypeptide expression (Gottesman, S., *Gene Expression Technology: Methods in Enzymology, Academic Press, San Diego, California 185*: 119-128 (1990)). Another strategy is to alter the nucleotide sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada *et al.*, *Nucleic Acids Res. 20*: 2111-2118 (1992)). Such alteration of nucleotide sequences can be carried out by standard DNA synthesis techniques.

[0049] When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. Recombinant mammalian expression vectors are



often capable of directing expression of the nucleic acid in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Non-limiting examples of suitable tissue-specific promoters include an albumin promoter (liver-specific; Pinkert *et al.*, *Genes Dev.* 1: 268-277 (1987)), lymphoid-specific promoters (Calame & Eaton, *Adv. Immunol.* 43: 235-275 (1988)), promoters of T cell receptors (Winoto & Baltimore, *EMBO J.* 8: 729-733 (1989)) promoters of immunoglobulins (Banerji *et al.*, *Cell* 33: 729-740 (1983); Queen & Baltimore, *Cell* 33: 741-748 (1983)), neuron-specific promoters (e.g., the neurofilament promoter; Byrne & Ruddle, *Proc. Natl. Acad. Sci. USA* 86: 5473-5477 (1989)), pancreas-specific promoters (Edlund *et al.*, *Science* 230: 912-916 (1985)), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are sometimes utilized, for example, the murine hox promoters (Kessel & Gruss, *Science* 249: 374-379 (1990)) and the  $\alpha$ -fetoprotein promoter (Campes & Tilghman, *Genes Dev.* 3: 537-546 (1989)).

**[0050]** A *ADAMTS2* nucleic acid also may be cloned into an expression vector in an antisense orientation. Regulatory sequences (e.g., viral promoters and/or enhancers) operatively linked to a *ADAMTS2* nucleic acid cloned in the antisense orientation can be chosen for directing constitutive, tissue specific or cell type specific expression of antisense RNA in a variety of cell types. Antisense expression vectors can be in the form of a recombinant plasmid, phagemid or attenuated virus. For a discussion of the regulation of gene expression using antisense genes see, e.g., Weintraub *et al.*, Antisense RNA as a molecular tool for genetic analysis, *Reviews - Trends in Genetics*, Vol. 1(1) (1986).

**[0051]** Also provided herein are host cells that include a *ADAMTS2* nucleotide sequence within a recombinant expression vector or a fragment of such a nucleotide sequence which facilitate homologous recombination into a specific site of the host cell genome. The terms "host cell" and "recombinant host cell" are used interchangeably herein. Such terms refer not only to the particular subject cell but rather also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein. A host cell can be any prokaryotic or eukaryotic cell. For example, a target polypeptide can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

**[0052]** Vectors can be introduced into host cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (e.g., DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, transduction/infection, DEAE-dextran-mediated transfection, lipofection, or electroporation.

[0053] A host cell provided herein can be used to produce (*i.e.*, express) a target polypeptide or a substantially identical polypeptide thereof. Accordingly, further provided are methods for producing a target polypeptide using host cells described herein. In one embodiment, the method includes culturing host cells into which a recombinant expression vector encoding a target polypeptide has been introduced in a suitable medium such that a target polypeptide is produced. In another embodiment, the method further includes isolating a target polypeptide from the medium or the host cell.

[0054] Also provided are cells or purified preparations of cells which include a *ADAMTS2* transgene, or which otherwise misexpress target polypeptide. Cell preparations can consist of human or non-human cells, *e.g.*, rodent cells, *e.g.*, mouse or rat cells, rabbit cells, or pig cells. In preferred embodiments, the cell or cells include a *ADAMTS2* transgene (*e.g.*, a heterologous form of a *ADAMTS2* gene, such as a human gene expressed in non-human cells). The transgene can be misexpressed, *e.g.*, overexpressed or underexpressed. In other preferred embodiments, the cell or cells include a gene which misexpress an endogenous target polypeptide (*e.g.*, expression of a gene is disrupted, also known as a knockout). Such cells can serve as a model for studying disorders which are related to mutated or mis-expressed alleles or for use in drug screening. Also provided are human cells (*e.g.*, a hematopoietic stem cells) transfected with a *ADAMTS2* nucleic acid.

[0055] Also provided are cells or a purified preparation thereof (*e.g.*, human cells) in which an endogenous *ADAMTS2* nucleic acid is under the control of a regulatory sequence that does not normally control the expression of the endogenous gene. The expression characteristics of an endogenous gene within a cell (*e.g.*, a cell line or microorganism) can be modified by inserting a heterologous DNA regulatory element into the genome of the cell such that the inserted regulatory element is operably linked to the corresponding endogenous gene. For example, an endogenous corresponding gene (*e.g.*, a gene which is “transcriptionally silent,” not normally expressed, or expressed only at very low levels) may be activated by inserting a regulatory element which is capable of promoting the expression of a normally expressed gene product in that cell. Techniques such as targeted homologous recombinations, can be used to insert the heterologous DNA as described in, *e.g.*, Chappel, US 5,272,071; WO 91/06667, published on May 16, 1991.

#### Transgenic Animals

[0056] Non-human transgenic animals that express a heterologous target polypeptide (*e.g.*, expressed from a *ADAMTS2* nucleic acid or substantially identical sequence thereof) can be generated. Such animals are useful for studying the function and/or activity of a target polypeptide and for identifying and/or evaluating modulators of the activity of *ADAMTS2* nucleic acids and encoded polypeptides. As used herein, a “transgenic animal” is a non-human animal such as a mammal (*e.g.*, a non-human primate such as chimpanzee, baboon, or macaque; an ungulate such as an equine, bovine, or

caprine; or a rodent such as a rat, a mouse, or an Israeli sand rat), a bird (*e.g.*, a chicken or a turkey), an amphibian (*e.g.*, a frog, salamander, or newt), or an insect (*e.g.*, *Drosophila melanogaster*), in which one or more of the cells of the animal includes a transgene. A transgene is exogenous DNA or a rearrangement (*e.g.*, a deletion of endogenous chromosomal DNA) that is often integrated into or occurs in the genome of cells in a transgenic animal. A transgene can direct expression of an encoded gene product in one or more cell types or tissues of the transgenic animal, and other transgenes can reduce expression (*e.g.*, a knockout). Thus, a transgenic animal can be one in which an endogenous nucleic acid homologous to a *ADAMTS2* nucleic acid has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal (*e.g.*, an embryonic cell of the animal) prior to development of the animal.

[0057] Intronic sequences and polyadenylation signals can also be included in the transgene to increase expression efficiency of the transgene. One or more tissue-specific regulatory sequences can be operably linked to a *ADAMTS2* nucleotide sequence to direct expression of an encoded polypeptide to particular cells. A transgenic founder animal can be identified based upon the presence of a *ADAMTS2* nucleotide sequence in its genome and/or expression of encoded mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a *ADAMTS2* nucleotide sequence can further be bred to other transgenic animals carrying other transgenes.

[0058] Target polypeptides can be expressed in transgenic animals or plants by introducing, for example, a *ADAMTS2* nucleic acid into the genome of an animal that encodes the target polypeptide. In preferred embodiments the nucleic acid is placed under the control of a tissue specific promoter, *e.g.*, a milk or egg specific promoter, and recovered from the milk or eggs produced by the animal. Also included is a population of cells from a transgenic animal.

#### Target Polypeptides

[0059] Also featured herein are isolated target polypeptides, which are encoded by a *ADAMTS2* nucleotide sequence (*e.g.*, SEQ ID NO: 1-3), or a substantially identical nucleotide sequence thereof. *ADAMTS2* exists in two forms, a "long" form comprising a molecule approximately 130 kDa in length (*e.g.*, SEQ ID NO: 4), and a "short" form comprising a molecule approximately 70 kDa in length (*e.g.*, SEQ ID NO: 5). Thus featured herein are isolated *ADAMTS2* polypeptides, which include long and short isoforms, and substantially identical polypeptides thereof. An *ADAMTS2* polypeptide is a polypeptide encoded by an *ADAMTS2* nucleic acid, where one nucleic acid can encode one or more different polypeptides. The term "polypeptide" as used herein includes proteins and peptides. An "isolated" or "purified" polypeptide or protein is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the protein is derived, or substantially free from chemical

precursors or other chemicals when chemically synthesized. In one embodiment, the language “substantially free” means preparation of a target polypeptide having less than about 30%, 20%, 10% and more preferably 5% (by dry weight), of non-target polypeptide (also referred to herein as a “contaminating protein”), or of chemical precursors or non-target chemicals. When the target polypeptide or a biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, specifically, where culture medium represents less than about 20%, sometimes less than about 10%, and often less than about 5% of the volume of the polypeptide preparation. Isolated or purified target polypeptide preparations are sometimes 0.01 milligrams or more or 0.1 milligrams or more, and often 1.0 milligrams or more and 10 milligrams or more in dry weight.

[0060] Further included herein are target polypeptide fragments. The polypeptide fragment may be a domain or part of a domain of a target polypeptide. The polypeptide fragment may have increased, decreased or unexpected biological activity. The polypeptide fragment is often 50 or fewer, 100 or fewer, or 200 or fewer amino acids in length, and is sometimes 300, 400, 500, 600, 700, or 900 or fewer amino acids in length. Thus, featured herein are *ADAMTS2* polypeptides and biologically active or antigenic fragments thereof useful as reagents or targets in assays applicable to treatment and diagnosis of osteoarthritis. In another embodiment, provided herein are *ADAMTS2* polypeptides having a *ADAMTS2* activity (e.g., a zinc binding activity, a metalloprotease activity, a procollagen II processing or synthesis activity, or a collagen II synthesis activity in vitro or in vivo). In certain embodiments, the polypeptides are *ADAMTS2* proteins including at least one propeptide domain, at least one metalloproteinase domain, at least one disintegrin-like domain, at least one, two, three, and often four thrombospondin domains, and sometimes having a *ADAMTS2* activity, e.g., a *ADAMTS2* activity as described herein. *ADAMTS2* polypeptides and fragments thereof often have biological activity, such as excising the N-propeptide of type II procollagens. Methods for monitoring and quantifying this biological activity are known (e.g., Colige et al., J. Biol. Chem. 270: 16724-16730 (1995)).

[0061] Human *ADAMTS2* protein (SEQ ID NO: 4-5) includes a signal sequence of about 29 amino acids (from amino acid 1 to about amino acid 29 of SEQ ID NO: 4-5). The *ADAMTS2* protein without the signal sequence can be approximately 1182 amino acid residues in length (from about amino acid 30 to amino acid 1211 of SEQ ID NO: 4) or approximately 485 amino acid residues in length (from about amino acid 30 to amino acid 514 of SEQ ID NO: 5). Human *ADAMTS2* protein includes a “pro” region homologous to the reprotin family propeptide, which is typically post-translationally cleaved upon conversion of the inactive (or pro-domain containing) protein to the catalytically active metalloprotease. The prodomain region of human *ADAMTS2* protein corresponds to about amino acids 30 to 251, 30 to 252, 30 to 253, 30 to 254, 30 to 255, 30 to 256, 30 to 257, 30 to 258 or 30 to 259 of SEQ ID NO: 4-5, where it is understood that the active form of *ADAMTS2* does not contain the propeptide domain.

[0062] Upon cleavage, catalytically active mature protein can be approximately 960, 959, 958, 957, 956, 955, 954, 953 or 952 amino acids in length (from about amino acid 252, 253, 254, 255, 256, 257, 258, 259 or 260 to amino acid 1211 of SEQ ID NO: 4) or approximately 261, 260, 259, 258, 257, 256, 255, 254 or 253 amino acid residues in length (from about amino acid 252, 253, 254, 255, 256, 257, 258, 259 or 260 to amino acid 514 of SEQ ID NO: 5).

[0063] Human *ADAMTS2* contains the following regions or other structural features: a signal sequence at about amino acids 1-29 of SEQ ID NO: 4-5; a reprotolysin family propeptide domain located at about amino acid residues 30 to 251, 30 to 252, 30 to 253, 30 to 254, 30 to 255, 30 to 256, 30 to 257, 30 to 258 or 30 to 259 of SEQ ID NO: 4-5; a zinc-metalloprotease catalytic domain at about amino acids 251 to 479, 252 to 479, 253 to 479, 254 to 479, 255 to 479, 256 to 479, 257 to 479, 258 to 479 or 259 to 479 of SEQ ID NO: 4-5; a disintegrin domain at about amino acids 480 to 560 of SEQ ID NO: 4; a cysteine-rich domain at about amino acids 618 to 722 of SEQ ID NO: 4; four thrombospondin motifs-2 motifs at about amino acids 561 to 616, 854 to 912, 914 to 971, and 975 to 1029 of SEQ ID NO: 4; and eight N-glycosylation sites located at about amino acids 112, 251, 949, 993, 1031, 1098, 1145, and 1150 of SEQ ID NO: 4.

[0064] In other embodiments, provided are methods of increasing the synthesis of procollagen II comprising providing or administering to individuals in need of increasing levels of type II collagen the pharmaceutical or physiologically acceptable composition comprising active human *ADAMTS2* protein or fragment thereof, where *ADAMTS2* polypeptide fragments having activity are selected from amino acids 252-1211, 253-1211, 254-1211, 255-1211, 256-1211, 257-1211, 258-1211, 259-1211 or 260-1211 of SEQ ID NO: 4, where it is understood that the active form of *ADAMTS2* does not contain the propeptide domain.

[0065] Substantially identical target polypeptides may depart from the amino acid sequences of target polypeptides in different manners. For example, conservative amino acid modifications may be introduced at one or more positions in the amino acid sequences of target polypeptides. A “conservative amino acid substitution” is one in which the amino acid is replaced by another amino acid having a similar structure and/or chemical function. Families of amino acid residues having similar structures and functions are well known. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Also, essential and non-essential amino acids may be replaced. A “non-essential” amino acid is one that can be altered without abolishing or substantially altering the biological function of a target polypeptide, whereas altering an “essential” amino acid abolishes or substantially alters the

biological function of a target polypeptide. Amino acids that are conserved among target polypeptides are typically essential amino acids. In certain embodiments, the polypeptide includes one or more non-synonymous polymorphic variants associated with osteoarthritis.

[0066] Also, target polypeptides may exist as chimeric or fusion polypeptides. As used herein, a target “chimeric polypeptide” or target “fusion polypeptide” includes a target polypeptide linked to a non-target polypeptide. A “non-target polypeptide” refers to a polypeptide having an amino acid sequence corresponding to a polypeptide which is not substantially identical to the target polypeptide, which includes, for example, a polypeptide that is different from the target polypeptide and derived from the same or a different organism. The target polypeptide in the fusion polypeptide can correspond to an entire or nearly entire target polypeptide or a fragment thereof. The non-target polypeptide can be fused to the N-terminus or C-terminus of the target polypeptide.

[0067] Fusion polypeptides can include a moiety having high affinity for a ligand. For example, the fusion polypeptide can be a GST-target fusion polypeptide in which the target sequences are fused to the C-terminus of the GST sequences, or a polyhistidine-target fusion polypeptide in which the target polypeptide is fused at the N- or C-terminus to a string of histidine residues. Such fusion polypeptides can facilitate purification of recombinant target polypeptide. Expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide), and a nucleotide sequence in SEQ ID NO: 1-3, or a substantially identical nucleotide sequence thereof, can be cloned into an expression vector such that the fusion moiety is linked in-frame to the target polypeptide. Further, the fusion polypeptide can be a target polypeptide containing a heterologous signal sequence at its N-terminus. In certain host cells (*e.g.*, mammalian host cells), expression, secretion, cellular internalization, and cellular localization of a target polypeptide can be increased through use of a heterologous signal sequence. Fusion polypeptides can also include all or a part of a serum polypeptide (*e.g.*, an IgG constant region or human serum albumin).

[0068] Target polypeptides can be incorporated into pharmaceutical compositions and administered to a subject *in vivo*. Administration of these target polypeptides can be used to affect the bioavailability of a substrate of the target polypeptide and may effectively increase target polypeptide biological activity in a cell. Target fusion polypeptides may be useful therapeutically for the treatment of disorders caused by, for example, (i) aberrant modification or mutation of a gene encoding a target polypeptide; (ii) misregulation of the gene encoding the target polypeptide; and (iii) aberrant post-translational modification of a target polypeptide. Also, target polypeptides can be used as immunogens to produce anti-target antibodies in a subject, to purify target polypeptide ligands or binding partners, and in screening assays to identify molecules which inhibit or enhance the interaction of a target polypeptide with a substrate.

[0069] In addition, polypeptides can be chemically synthesized using techniques known in the art (See, *e.g.*, Creighton, 1983 Proteins. New York, N.Y.: W. H. Freeman and Company; and Hunkapiller et

al., (1984) Nature July 12 -18;310(5973):105-11). For example, a relative short fragment can be synthesized by use of a peptide synthesizer. Furthermore, if desired, non-classical amino acids or chemical amino acid analogs can be introduced as a substitution or addition into the fragment sequence. Non-classical amino acids include, but are not limited to, to the D-isomers of the common amino acids, 2,4-diaminobutyric acid, α-amino isobutyric acid, 4-aminobutyric acid, Abu, 2-amino butyric acid, g-Abu, ε-Ahx, 6-amino hexanoic acid, Aib, 2-amino isobutyric acid, 3-amino propionic acid, ornithine, norleucine, norvaline, hydroxyproline, sarcosine, citrulline, homocitrulline, cysteic acid, t-butylglycine, t-butylalanine, phenylglycine, cyclohexylalanine, β-alanine, fluoroamino acids, designer amino acids such as β-methyl amino acids, Ca-methyl amino acids, Na-methyl amino acids, and amino acid analogs in general. Furthermore, the amino acid can be D (dextrorotary) or L (levorotary).

[0070] Polypeptides and polypeptide fragments sometimes are differentially modified during or after translation, *e.g.*, by glycosylation, acetylation, phosphorylation, amidation, derivatization by known protecting/blocking groups, proteolytic cleavage, linkage to an antibody molecule or other cellular ligand, etc. Any of numerous chemical modifications may be carried out by known techniques, including but not limited, to specific chemical cleavage by cyanogen bromide, trypsin, chymotrypsin, papain, V8 protease, NaBH<sub>4</sub>; acetylation, formylation, oxidation, reduction; metabolic synthesis in the presence of tunicamycin; and the like. Additional post-translational modifications include, for example, N-linked or O-linked carbohydrate chains, processing of N-terminal or C-terminal ends), attachment of chemical moieties to the amino acid backbone, chemical modifications of N-linked or O-linked carbohydrate chains, and addition or deletion of an N-terminal methionine residue as a result of prokaryotic host cell expression. The polypeptide fragments may also be modified with a detectable label, such as an enzymatic, fluorescent, isotopic or affinity label to allow for detection and isolation of the polypeptide.

[0071] Also provided are chemically modified derivatives of polypeptides that can provide additional advantages such as increased solubility, stability and circulating time of the polypeptide, or decreased immunogenicity (*see e.g.*, U.S. Pat. No: 4,179,337. The chemical moieties for derivitization may be selected from water soluble polymers such as polyethylene glycol, ethylene glycol/propylene glycol copolymers, carboxymethylcellulose, dextran, polyvinyl alcohol and the like. The polypeptides may be modified at random positions within the molecule, or at predetermined positions within the molecule and may include one, two, three or more attached chemical moieties.

[0072] The polymer may be of any molecular weight, and may be branched or unbranched. For polyethylene glycol, the preferred molecular weight is between about 1 kDa and about 100 kDa (the term "about" indicating that in preparations of polyethylene glycol, some molecules will weigh more, some less, than the stated molecular weight) for ease in handling and manufacturing. Other sizes may be used, depending on the desired therapeutic profile (*e.g.*, the duration of sustained release desired, the effects, if

any on biological activity, the ease in handling, the degree or lack of antigenicity and other known effects of the polyethylene glycol to a therapeutic protein or analog).

[0073] The polymers should be attached to the polypeptide with consideration of effects on functional or antigenic domains of the polypeptide. There are a number of attachment methods available to those skilled in the art (*e.g.*, EP 0 401 384 (coupling PEG to G-CSF) and Malik et al. (1992) *Exp Hematol.* September;20(8):1028-35 (pegylation of GM-CSF using tresyl chloride)). For example, polyethylene glycol may be covalently bound through amino acid residues via a reactive group, such as a free amino or carboxyl group. Reactive groups are those to which an activated polyethylene glycol molecule may be bound. The amino acid residues having a free amino group may include lysine residues and the N-terminal amino acid residues; those having a free carboxyl group may include aspartic acid residues, glutamic acid residues and the C-terminal amino acid residue. Sulfhydryl groups may also be used as a reactive group for attaching the polyethylene glycol molecules. For therapeutic purposes, the attachment sometimes is at an amino group, such as attachment at the N-terminus or lysine group.

[0074] Proteins can be chemically modified at the N-terminus. Using polyethylene glycol as an illustration of such a composition, one may select from a variety of polyethylene glycol molecules (by molecular weight, branching, and the like), the proportion of polyethylene glycol molecules to protein (polypeptide) molecules in the reaction mix, the type of pegylation reaction to be performed, and the method of obtaining the selected N-terminally pegylated protein. The method of obtaining the N-terminally pegylated preparation (*i.e.*, separating this moiety from other monopegylated moieties if necessary) may be by purification of the N-terminally pegylated material from a population of pegylated protein molecules. Selective proteins chemically modified at the N-terminus may be accomplished by reductive alkylation, which exploits differential reactivity of different types of primary amino groups (lysine versus the N-terminal) available for derivatization in a particular protein. Under the appropriate reaction conditions, substantially selective derivatization of the protein at the N-terminus with a carbonyl group containing polymer is achieved.

#### Substantially Identical Nucleic Acids and Polypeptides

[0075] Nucleotide sequences and polypeptide sequences that are substantially identical to a *ADAMTS2* nucleotide sequence and the target polypeptide sequences encoded by those nucleotide sequences, respectively, are included herein. The term "substantially identical" as used herein refers to two or more nucleic acids or polypeptides sharing one or more identical nucleotide sequences or polypeptide sequences, respectively. Included are nucleotide sequences or polypeptide sequences that are 55% or more, 60% or more, 65% or more, 70% or more, 75% or more, 80% or more, 85% or more, 90% or more, 95% or more (each often within a 1%, 2%, 3% or 4% variability) identical to a *ADAMTS2* nucleotide sequence or the encoded target polypeptide amino acid sequences. One test for determining



whether two nucleic acids are substantially identical is to determine the percent of identical nucleotide sequences or polypeptide sequences shared between the nucleic acids or polypeptides.

[0076] Calculations of sequence identity are often performed as follows. Sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). The length of a reference sequence aligned for comparison purposes is sometimes 30% or more, 40% or more, 50% or more, often 60% or more, and more often 70% or more, 80% or more, 90% or more, or 100% of the length of the reference sequence. The nucleotides or amino acids at corresponding nucleotide or polypeptide positions, respectively, are then compared among the two sequences. When a position in the first sequence is occupied by the same nucleotide or amino acid as the corresponding position in the second sequence, the nucleotides or amino acids are deemed to be identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, introduced for optimal alignment of the two sequences.

[0077] Comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. Percent identity between two amino acid or nucleotide sequences can be determined using the algorithm of Meyers & Miller, *CABIOS* 4: 11-17 (1989), which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4. Also, percent identity between two amino acid sequences can be determined using the Needleman & Wunsch, *J. Mol. Biol.* 48: 444-453 (1970) algorithm which has been incorporated into the GAP program in the GCG software package (available at the [http](http://www.gcg.com) address [www.gcg.com](http://www.gcg.com)), using either a Blossum 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. Percent identity between two nucleotide sequences can be determined using the GAP program in the GCG software package (available at [http](http://www.gcg.com) address [www.gcg.com](http://www.gcg.com)), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. A set of parameters often used is a Blossum 62 scoring matrix with a gap open penalty of 12, a gap extend penalty of 4, and a frameshift gap penalty of 5.

[0078] Another manner for determining if two nucleic acids are substantially identical is to assess whether a polynucleotide homologous to one nucleic acid will hybridize to the other nucleic acid under stringent conditions. As used herein, the term "stringent conditions" refers to conditions for hybridization and washing. Stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y., 6.3.1-6.3.6 (1989). Aqueous and non-aqueous methods are described in that reference and either can be used. An example of stringent hybridization conditions is hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C,

followed by one or more washes in 0.2X SSC, 0.1% SDS at 50°C. Another example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 55°C. A further example of stringent hybridization conditions is hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 60°C. Often, stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 65°C. More often, stringency conditions are 0.5M sodium phosphate, 7% SDS at 65°C, followed by one or more washes at 0.2X SSC, 1% SDS at 65°C.

[0079] An example of a substantially identical nucleotide sequence to a nucleotide sequence in SEQ ID NO: 1-3 is one that has a different nucleotide sequence but still encodes the same polypeptide sequence encoded by the nucleotide sequence in SEQ ID NO: 1-3. Another example is a nucleotide sequence that encodes a polypeptide having a polypeptide sequence that is more than 70% or more identical to, sometimes more than 75% or more, 80% or more, or 85% or more identical to, and often more than 90% or more and 95% or more identical to a polypeptide sequence encoded by a nucleotide sequence in SEQ ID NO: 1-3.

[0080] Nucleotide sequences in SEQ ID NO: 1-3 and amino acid sequences of encoded polypeptides can be used as “query sequences” to perform a search against public databases to identify other family members or related sequences, for example. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul *et al.*, *J. Mol. Biol.* 215: 403-10 (1990). BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to nucleotide sequences in SEQ ID NO: 1-3. BLAST polypeptide searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to polypeptides encoded by the nucleotide sequences of SEQ ID NO: 1-3. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.*, *Nucleic Acids Res.* 25(17): 3389-3402 (1997). When utilizing BLAST and Gapped BLAST programs, default parameters of the respective programs (*e.g.*, XBLAST and NBLAST) can be used (*see* the http address [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)).

[0081] A nucleic acid that is substantially identical to a nucleotide sequence in SEQ ID NO: 1-3 may include polymorphic sites at positions equivalent to those described herein when the sequences are aligned. For example, using the alignment procedures described herein, SNPs in a sequence substantially identical to a sequence in SEQ ID NO: 1-3 can be identified at nucleotide positions that match (*i.e.*, align) with nucleotides at SNP positions in each nucleotide sequence in SEQ ID NO: 1-3. Also, where a polymorphic variation results in an insertion or deletion, insertion or deletion of a nucleotide sequence

from a reference sequence can change the relative positions of other polymorphic sites in the nucleotide sequence.

[0082] Substantially identical nucleotide and polypeptide sequences include those that are naturally occurring, such as allelic variants (same locus), splice variants, homologs (different locus), and orthologs (different organism) or can be non-naturally occurring. Non-naturally occurring variants can be generated by mutagenesis techniques, including those applied to polynucleotides, cells, or organisms. The variants can contain nucleotide substitutions, deletions, inversions and insertions. Variation can occur in either or both the coding and non-coding regions. The variations can produce both conservative and non-conservative amino acid substitutions (as compared in the encoded product). Orthologs, homologs, allelic variants, and splice variants can be identified using methods known in the art. These variants normally comprise a nucleotide sequence encoding a polypeptide that is 50% or more, about 55% or more, often about 70-75% or more or about 80-85% or more, and sometimes about 90-95% or more identical to the amino acid sequences of target polypeptides or a fragment thereof. Such nucleic acid molecules can readily be identified as being able to hybridize under stringent conditions to a nucleotide sequence in SEQ ID NO: 1-3 or a fragment of this sequence. Nucleic acid molecules corresponding to orthologs, homologs, and allelic variants of a nucleotide sequence in SEQ ID NO: 1-3 can further be identified by mapping the sequence to the same chromosome or locus as the nucleotide sequence in SEQ ID NO: 1-3.

[0083] Also, substantially identical nucleotide sequences may include codons that are altered with respect to the naturally occurring sequence for enhancing expression of a target polypeptide in a particular expression system. For example, the nucleic acid can be one in which one or more codons are altered, and often 10% or more or 20% or more of the codons are altered for optimized expression in bacteria (*e.g.*, *E. coli.*), yeast (*e.g.*, *S. cerevisiae*), human (*e.g.*, 293 cells), insect, or rodent (*e.g.*, hamster) cells.

#### Methods for Identifying Risk of Osteoarthritis

[0084] Methods for prognosing and diagnosing osteoarthritis are included herein. These methods include detecting the presence or absence of one or more polymorphic variations in a nucleotide sequence associated with osteoarthritis, such as variants in or around the loci set forth herein, or a substantially identical sequence thereof, in a sample from a subject, where the presence of a polymorphic variant described herein is indicative of a risk of osteoarthritis. Determining a risk of osteoarthritis sometimes refers to determining whether an individual is at an increased risk of osteoarthritis (*e.g.*, intermediate risk or higher risk).

[0085] Thus, featured herein is a method for identifying a subject who is at risk of osteoarthritis, which comprises detecting an aberration associated with osteoarthritis in a nucleic acid sample from the

subject. An embodiment is a method for detecting a risk of osteoarthritis in a subject, which comprises detecting the presence or absence of a polymorphic variation associated with osteoarthritis at a polymorphic site in a nucleotide sequence in a nucleic acid sample from a subject, where the nucleotide sequence comprises a polynucleotide sequence selected from the group consisting of: (a) a nucleotide sequence of SEQ ID NO: 1-3; (b) a nucleotide sequence which encodes a polypeptide consisting of an amino acid sequence encoded by a nucleotide sequence of SEQ ID NO: 1-3; (c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to an amino acid sequence encoded by a nucleotide sequence of SEQ ID NO: 1-3, or a nucleotide sequence about 90% or more identical to a nucleotide sequence of SEQ ID NO: 1-3; and (d) a fragment of a nucleotide sequence of (a), (b), or (c) comprising the polymorphic site; whereby the presence of the polymorphic variation is indicative of a predisposition to osteoarthritis in the subject. In some embodiments, a polymorphic variation at position 733 of SEQ ID NO: 2-3 may be detected (e.g., a guanine in these sequences or a cytosine in a complementary sequence are associated with increased risk of osteoarthritis). In certain embodiments, polymorphic variants at the positions described herein are detected for determining a risk of osteoarthritis, and polymorphic variants at positions in linkage disequilibrium with these positions are detected for determining a risk of osteoarthritis. As used herein, "SEQ ID NO: 1-3" refers to individual sequences in SEQ ID NO: 1, 2 or 3, each sequence being separately applicable to embodiments described herein.

[0086] Risk of osteoarthritis sometimes is expressed as a probability, such as an odds ratio, percentage, or risk factor. Risk often is based upon the presence or absence of one or more polymorphic variants described herein, and also may be based in part upon phenotypic traits of the individual being tested. Methods for calculating risk based upon patient data are well known (*see, e.g.*, Agresti, *Categorical Data Analysis*, 2nd Ed. 2002. Wiley). Allelotyping and genotyping analyses may be carried out in populations other than those exemplified herein to enhance the predictive power of the prognostic method. These further analyses are executed in view of the exemplified procedures described herein, and may be based upon the same polymorphic variations or additional polymorphic variations.

[0087] In certain embodiments, determining the presence of a combination of two or more polymorphic variants associated with osteoarthritis in one or more genetic loci (e.g., one or more genes) of the sample is determined to identify, quantify and/or estimate, risk of osteoarthritis. The risk often is the probability of having or developing osteoarthritis. The risk sometimes is expressed as a relative risk with respect to a population average risk of osteoarthritis, and sometimes is expressed as a relative risk with respect to the lowest risk group. Such relative risk assessments often are based upon penetrance values determined by statistical methods, and are particularly useful to clinicians and insurance companies for assessing risk of osteoarthritis (e.g., a clinician can target appropriate detection, prevention and therapeutic regimens to a patient after determining the patient's risk of osteoarthritis, and an

insurance company can fine tune actuarial tables based upon population genotype assessments of osteoarthritis risk). Risk of osteoarthritis sometimes is expressed as an odds ratio, which is the odds of a particular person having a genotype has or will develop osteoarthritis with respect to another genotype group (e.g., the most disease protective genotype or population average). In related embodiments, the determination is utilized to identify a subject at risk of osteoarthritis. In an embodiment, two or more polymorphic variations are detected in two or more regions in human genomic DNA associated with increased risk of osteoarthritis, such as a locus containing a *ADAMTS2*, for example. In certain embodiments, 3 or more, or 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20 or more polymorphic variants are detected in the sample. In specific embodiments, polymorphic variants are detected in a *ADAMTS2* region, for example. In certain embodiments, polymorphic variants are detected at other genetic loci (e.g., the polymorphic variants can be detected in *ADAMTS2* in addition to other loci or only in other loci), where the other loci include but are not limited to those described in concurrently-filed patent applications having attorney docket number 524593008700, 524593008800, 524593009000 or 524593009200, which is incorporated herein by reference in its entirety.

[0088] Results from prognostic tests may be combined with other test results to diagnose osteoarthritis. For example, prognostic results may be gathered, a patient sample may be ordered based on a determined predisposition to osteoarthritis, the patient sample is analyzed, and the results of the analysis may be utilized to diagnose osteoarthritis. Also osteoarthritis diagnostic method can be developed from studies used to generate prognostic methods in which populations are stratified into subpopulations having different progressions of osteoarthritis. In another embodiment, prognostic results may be gathered, a patient's risk factors for developing osteoarthritis (e.g., age, weight, race, diet) analyzed, and a patient sample may be ordered based on a determined predisposition to osteoarthritis.

[0089] The nucleic acid sample typically is isolated from a biological sample obtained from a subject. For example, nucleic acid can be isolated from blood, saliva, sputum, urine, cell scrapings, and biopsy tissue. The nucleic acid sample can be isolated from a biological sample using standard techniques, such as the technique described in Example 2. As used herein, the term "subject" refers primarily to humans but also refers to other mammals such as dogs, cats, and ungulates (e.g., cattle, sheep, and swine). Subjects also include avians (e.g., chickens and turkeys), reptiles, and fish (e.g., salmon), as embodiments described herein can be adapted to nucleic acid samples isolated from any of these organisms. The nucleic acid sample may be isolated from the subject and then directly utilized in a method for determining the presence of a polymorphic variant, or alternatively, the sample may be isolated and then stored (e.g., frozen) for a period of time before being subjected to analysis.

[0090] The presence or absence of a polymorphic variant is determined using one or both chromosomal complements represented in the nucleic acid sample. Determining the presence or absence of a polymorphic variant in both chromosomal complements represented in a nucleic acid sample from a

subject having a copy of each chromosome is useful for determining the zygosity of an individual for the polymorphic variant (*i.e.*, whether the individual is homozygous or heterozygous for the polymorphic variant). Any oligonucleotide-based diagnostic may be utilized to determine whether a sample includes the presence or absence of a polymorphic variant in a sample. For example, primer extension methods, ligase sequence determination methods (*e.g.*, U.S. Pat. Nos. 5,679,524 and 5,952,174, and WO 01/27326), mismatch sequence determination methods (*e.g.*, U.S. Pat. Nos. 5,851,770; 5,958,692; 6,110,684; and 6,183,958), microarray sequence determination methods, restriction fragment length polymorphism (RFLP), single strand conformation polymorphism detection (SSCP) (*e.g.*, U.S. Pat. Nos. 5,891,625 and 6,013,499), PCR-based assays (*e.g.*, TAQMAN<sup>®</sup> PCR System (Applied Biosystems)), and nucleotide sequencing methods may be used.

[0091] Oligonucleotide extension methods typically involve providing a pair of oligonucleotide primers in a polymerase chain reaction (PCR) or in other nucleic acid amplification methods for the purpose of amplifying a region from the nucleic acid sample that comprises the polymorphic variation. One oligonucleotide primer is complementary to a region 3' of the polymorphism and the other is complementary to a region 5' of the polymorphism. A PCR primer pair may be used in methods disclosed in U.S. Pat. Nos. 4,683,195; 4,683,202; 4,965,188; 5,656,493; 5,998,143; 6,140,054; WO 01/27327; and WO 01/27329 for example. PCR primer pairs may also be used in any commercially available machines that perform PCR, such as any of the GENEAMP<sup>®</sup> Systems available from Applied Biosystems. Also, those of ordinary skill in the art will be able to design oligonucleotide primers based upon a *ADAMTS2* nucleotide sequence using knowledge available in the art.

[0092] Also provided is an extension oligonucleotide that hybridizes to the amplified fragment adjacent to the polymorphic variation. As used herein, the term "adjacent" refers to the 3' end of the extension oligonucleotide being often 1 nucleotide from the 5' end of the polymorphic site, and sometimes 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides from the 5' end of the polymorphic site, in the nucleic acid when the extension oligonucleotide is hybridized to the nucleic acid. The extension oligonucleotide then is extended by one or more nucleotides, and the number and/or type of nucleotides that are added to the extension oligonucleotide determine whether the polymorphic variant is present. Oligonucleotide extension methods are disclosed, for example, in U.S. Pat. Nos. 4,656,127; 4,851,331; 5,679,524; 5,834,189; 5,876,934; 5,908,755; 5,912,118; 5,976,802; 5,981,186; 6,004,744; 6,013,431; 6,017,702; 6,046,005; 6,087,095; 6,210,891; and WO 01/20039. Oligonucleotide extension methods using mass spectrometry are described, for example, in U.S. Pat. Nos. 5,547,835; 5,605,798; 5,691,141; 5,849,542; 5,869,242; 5,928,906; 6,043,031; and 6,194,144, and a method often utilized is described herein in Example 2.

[0093] A microarray can be utilized for determining whether a polymorphic variant is present or absent in a nucleic acid sample. A microarray may include any oligonucleotides described herein, and

methods for making and using oligonucleotide microarrays suitable for diagnostic use are disclosed in U.S. Pat. Nos. 5,492,806; 5,525,464; 5,589,330; 5,695,940; 5,849,483; 6,018,041; 6,045,996; 6,136,541; 6,142,681; 6,156,501; 6,197,506; 6,223,127; 6,225,625; 6,229,911; 6,239,273; WO 00/52625; WO 01/25485; and WO 01/29259. The microarray typically comprises a solid support and the oligonucleotides may be linked to this solid support by covalent bonds or by non-covalent interactions. The oligonucleotides may also be linked to the solid support directly or by a spacer molecule. A microarray may comprise one or more oligonucleotides complementary to a polymorphic site set forth herein.

[0094] A kit also may be utilized for determining whether a polymorphic variant is present or absent in a nucleic acid sample. A kit often comprises one or more pairs of oligonucleotide primers useful for amplifying a fragment of a nucleotide sequence of SEQ ID NO: 1-3 or a substantially identical sequence thereof, where the fragment includes a polymorphic site. The kit sometimes comprises a polymerizing agent, for example, a thermostable nucleic acid polymerase such as one disclosed in U.S. Pat. Nos. 4,889,818 or 6,077,664. Also, the kit often comprises an elongation oligonucleotide that hybridizes to a *ADAMTS2* nucleotide sequence in a nucleic acid sample adjacent to the polymorphic site. Where the kit includes an elongation oligonucleotide, it also often comprises chain elongating nucleotides, such as dATP, dTTP, dGTP, dCTP, and dITP, including analogs of dATP, dTTP, dGTP, dCTP and dITP, provided that such analogs are substrates for a thermostable nucleic acid polymerase and can be incorporated into a nucleic acid chain elongated from the extension oligonucleotide. Along with chain elongating nucleotides would be one or more chain terminating nucleotides such as ddATP, ddTTP, ddGTP, ddCTP, and the like. In an embodiment, the kit comprises one or more oligonucleotide primer pairs, a polymerizing agent, chain elongating nucleotides, at least one elongation oligonucleotide, and one or more chain terminating nucleotides. Kits optionally include buffers, vials, microtiter plates, and instructions for use.

[0095] An individual identified as being at risk of osteoarthritis may be heterozygous or homozygous with respect to the allele associated with a higher risk of osteoarthritis. A subject homozygous for an allele associated with an increased risk of osteoarthritis is at a comparatively high risk of osteoarthritis, a subject heterozygous for an allele associated with an increased risk of osteoarthritis is at a comparatively intermediate risk of osteoarthritis, and a subject homozygous for an allele associated with a decreased risk of osteoarthritis is at a comparatively low risk of osteoarthritis. A genotype may be assessed for a complementary strand, such that the complementary nucleotide at a particular position is detected.

[0096] Also featured are methods for determining risk of osteoarthritis and/or identifying a subject at risk of osteoarthritis by contacting a polypeptide or protein encoded by a *ADAMTS2* nucleotide sequence from a subject with an antibody that specifically binds to an epitope associated with increased

risk of osteoarthritis in the polypeptide (e.g., an epitope comprising a valine at position 245 in an *IRILRI* polypeptide).

Applications of Prognostic and Diagnostic Results to Pharmacogenomic Methods

[0097] Pharmacogenomics is a discipline that involves tailoring a treatment for a subject according to the subject's genotype as a particular treatment regimen may exert a differential effect depending upon the subject's genotype. For example, based upon the outcome of a prognostic test described herein, a clinician or physician may target pertinent information and preventative or therapeutic treatments to a subject who would be benefited by the information or treatment and avoid directing such information and treatments to a subject who would not be benefited (e.g., the treatment has no therapeutic effect and/or the subject experiences adverse side effects).

[0098] The following is an example of a pharmacogenomic embodiment. A particular treatment regimen can exert a differential effect depending upon the subject's genotype. Where a candidate therapeutic exhibits a significant interaction with a major allele and a comparatively weak interaction with a minor allele (e.g., an order of magnitude or greater difference in the interaction), such a therapeutic typically would not be administered to a subject genotyped as being homozygous for the minor allele, and sometimes not administered to a subject genotyped as being heterozygous for the minor allele. In another example, where a candidate therapeutic is not significantly toxic when administered to subjects who are homozygous for a major allele but is comparatively toxic when administered to subjects heterozygous or homozygous for a minor allele, the candidate therapeutic is not typically administered to subjects who are genotyped as being heterozygous or homozygous with respect to the minor allele.

[0099] The methods described herein are applicable to pharmacogenomic methods for preventing, alleviating or treating osteoarthritis. For example, a nucleic acid sample from an individual may be subjected to a prognostic test described herein. Where one or more polymorphic variations associated with increased risk of osteoarthritis are identified in a subject, information for preventing or treating osteoarthritis and/or one or more osteoarthritis treatment regimens then may be prescribed to that subject.

[0100] In certain embodiments, a treatment or preventative regimen is specifically prescribed and/or administered to individuals who will most benefit from it based upon their risk of developing osteoarthritis assessed by the methods described herein. Thus, provided are methods for identifying a subject predisposed to osteoarthritis and then prescribing a therapeutic or preventative regimen to individuals identified as having a predisposition. Thus, certain embodiments are directed to a method for reducing osteoarthritis in a subject, which comprises: detecting the presence or absence of a polymorphic variant associated with osteoarthritis in a nucleotide sequence in a nucleic acid sample from a subject, where the nucleotide sequence comprises a polynucleotide sequence selected from the group consisting of: (a) a nucleotide sequence of SEQ ID NO: 1-3; (b) a nucleotide sequence which encodes a polypeptide



consisting of an amino acid sequence encoded by a nucleotide sequence of SEQ ID NO: 1-3; (c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to an amino acid sequence encoded by a nucleotide sequence of SEQ ID NO: 1-3, or a nucleotide sequence about 90% or more identical to a nucleotide sequence of SEQ ID NO: 1-3; and (d) a fragment of a polynucleotide sequence of (a), (b), or (c); and prescribing or administering a treatment regimen to a subject from whom the sample originated where the presence of a polymorphic variation associated with osteoarthritis is detected in the nucleotide sequence. In these methods, predisposition results may be utilized in combination with other test results to diagnose osteoarthritis.

**[0101]** Certain preventative treatments often are prescribed to subjects having a predisposition to osteoarthritis and where the subject is diagnosed with osteoarthritis or is diagnosed as having symptoms indicative of an early stage of osteoarthritis. The treatment sometimes is preventative (*e.g.*, is prescribed or administered to reduce the probability that osteoarthritis arises or progresses), sometimes is therapeutic, and sometimes delays, alleviates or halts the progression of osteoarthritis. Any known preventative or therapeutic treatment for alleviating or preventing the occurrence of osteoarthritis is prescribed and/or administered. For example, the treatment often is directed to decreasing pain and improving joint movement. Examples of OA treatments include exercises to keep joints flexible and improve muscle strength. Different medications to control pain, including corticosteroids and nonsteroidal anti-inflammatory drugs (NSAIDs, *e.g.*, Voltaren); cyclooxygenase-2 (COX-2) inhibitors (*e.g.*, Celebrex, Vioxx, Mobic, and Bextra); monoclonal antibodies (*e.g.*, Remicade); tumor necrosis factor inhibitors (*e.g.*, Enbrel); or injections of glucocorticoids, hyaluronic acid or chondroitin sulfate into joints that are inflamed and not responsive to NSAIDs. Orally administered chondroitin sulfate also may be used as a therapeutic, as it may increase hyaluronic acid levels and viscosity of synovial fluid, and decrease collagenase levels in synovial fluid. Also, glucosamine can serve as an OA therapeutic as delivering it into joints may inhibit enzymes involved in cartilage degradation and enhance the production of hyaluronic acid. For mild pain without inflammation, acetaminophen may be used. Other treatments include: heat/cold therapy for temporary pain relief; joint protection to prevent strain or stress on painful joints; surgery to relieve chronic pain in damaged joints; and weight control to prevent extra stress on weight-bearing joints.

**[0102]** As therapeutic approaches for treating osteoarthritis continue to evolve and improve, the goal of treatments for osteoarthritis related disorders is to intervene even before clinical signs first manifest. Thus, genetic markers associated with susceptibility to osteoarthritis prove useful for early diagnosis, prevention and treatment of osteoarthritis.

**[0103]** As osteoarthritis preventative and treatment information can be specifically targeted to subjects in need thereof (*e.g.*, those at risk of developing osteoarthritis or those in an early stage of osteoarthritis), provided herein is a method for preventing or reducing the risk of developing

osteoarthritis in a subject, which comprises: (a) detecting the presence or absence of a polymorphic variation associated with osteoarthritis at a polymorphic site in a nucleotide sequence in a nucleic acid sample from a subject; (b) identifying a subject with a predisposition to osteoarthritis, whereby the presence of the polymorphic variation is indicative of a predisposition to osteoarthritis in the subject; and (c) if such a predisposition is identified, providing the subject with information about methods or products to prevent or reduce osteoarthritis or to delay the onset of osteoarthritis. Also provided is a method of targeting information or advertising to a subpopulation of a human population based on the subpopulation being genetically predisposed to a disease or condition, which comprises: (a) detecting the presence or absence of a polymorphic variation associated with osteoarthritis at a polymorphic site in a nucleotide sequence in a nucleic acid sample from a subject; (b) identifying the subpopulation of subjects in which the polymorphic variation is associated with osteoarthritis; and (c) providing information only to the subpopulation of subjects about a particular product which may be obtained and consumed or applied by the subject to help prevent or delay onset of the disease or condition.

**[0104]** Pharmacogenomics methods also may be used to analyze and predict a response to osteoarthritis treatment or a drug. For example, if pharmacogenomics analysis indicates a likelihood that an individual will respond positively to osteoarthritis treatment with a particular drug, the drug may be administered to the individual. Conversely, if the analysis indicates that an individual is likely to respond negatively to treatment with a particular drug, an alternative course of treatment may be prescribed. A negative response may be defined as either the absence of an efficacious response or the presence of toxic side effects. The response to a therapeutic treatment can be predicted in a background study in which subjects in any of the following populations are genotyped: a population that responds favorably to a treatment regimen, a population that does not respond significantly to a treatment regimen, and a population that responds adversely to a treatment regimen (*e.g.*, exhibits one or more side effects). These populations are provided as examples and other populations and subpopulations may be analyzed. Based upon the results of these analyses, a subject is genotyped to predict whether he or she will respond favorably to a treatment regimen, not respond significantly to a treatment regimen, or respond adversely to a treatment regimen.

**[0105]** The tests described herein also are applicable to clinical drug trials. One or more polymorphic variants indicative of response to an agent for treating osteoarthritis or to side effects to an agent for treating osteoarthritis may be identified using the methods described herein. Thereafter, potential participants in clinical trials of such an agent may be screened to identify those individuals most likely to respond favorably to the drug and exclude those likely to experience side effects. In that way, the effectiveness of drug treatment may be measured in individuals who respond positively to the drug, without lowering the measurement as a result of the inclusion of individuals who are unlikely to respond positively in the study and without risking undesirable safety problems.

[0106] Thus, another embodiment is a method of selecting an individual for inclusion in a clinical trial of a treatment or drug comprising the steps of: (a) obtaining a nucleic acid sample from an individual; (b) determining the identity of a polymorphic variation which is associated with a positive response to the treatment or the drug, or at least one polymorphic variation which is associated with a negative response to the treatment or the drug in the nucleic acid sample, and (c) including the individual in the clinical trial if the nucleic acid sample contains said polymorphic variation associated with a positive response to the treatment or the drug or if the nucleic acid sample lacks said polymorphic variation associated with a negative response to the treatment or the drug. In addition, the methods described herein for selecting an individual for inclusion in a clinical trial of a treatment or drug encompass methods with any further limitation described in this disclosure, or those following, specified alone or in any combination. The polymorphic variation may be in a sequence selected individually or in any combination from the group consisting of (i) a nucleotide sequence of SEQ ID NO: 1-3; (ii) a nucleotide sequence which encodes a polypeptide consisting of an amino acid sequence encoded by a nucleotide sequence of SEQ ID NO: 1-3; (iii) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to an amino acid sequence encoded by a nucleotide sequence of SEQ ID NO: 1-3, or a nucleotide sequence about 90% or more identical to a nucleotide sequence of SEQ ID NO: 1-3; and (iv) a fragment of a polynucleotide sequence of (i), (ii), or (iii) comprising the polymorphic site. The including step (c) optionally comprises administering the drug or the treatment to the individual if the nucleic acid sample contains the polymorphic variation associated with a positive response to the treatment or the drug and the nucleic acid sample lacks said biallelic marker associated with a negative response to the treatment or the drug.

[0107] Also provided herein is a method of partnering between a diagnostic/prognostic testing provider and a provider of a consumable product, which comprises: (a) the diagnostic/prognostic testing provider detects the presence or absence of a polymorphic variation associated with osteoarthritis at a polymorphic site in a nucleotide sequence in a nucleic acid sample from a subject; (b) the diagnostic/prognostic testing provider identifies the subpopulation of subjects in which the polymorphic variation is associated with osteoarthritis; (c) the diagnostic/prognostic testing provider forwards information to the subpopulation of subjects about a particular product which may be obtained and consumed or applied by the subject to help prevent or delay onset of the disease or condition; and (d) the provider of a consumable product forwards to the diagnostic test provider a fee every time the diagnostic/prognostic test provider forwards information to the subject as set forth in step (c) above.

#### Compositions Comprising Osteoarthritis-Directed Molecules

[0108] Featured herein is a composition comprising a cell from a subject having osteoarthritis or at risk of osteoarthritis and one or more molecules specifically directed and targeted to a nucleic acid

comprising a *ADAMTS2* nucleotide sequence or amino acid sequence. Such directed molecules include, but are not limited to, a compound that binds to a *ADAMTS2* nucleotide sequence or amino acid sequence referenced herein; a RNAi or siRNA molecule having a strand complementary or substantially complementary to a *ADAMTS2* nucleotide sequence (e.g., hybridizes to a *ADAMTS2* nucleotide sequence under conditions of high stringency); an antisense nucleic acid complementary or substantially complementary to an RNA encoded by a *ADAMTS2* nucleotide sequence (e.g., hybridizes to a *ADAMTS2* nucleotide sequence under conditions of high stringency); a ribozyme that hybridizes to a *ADAMTS2* nucleotide sequence (e.g., hybridizes to a *ADAMTS2* nucleotide sequence under conditions of high stringency); a nucleic acid aptamer that specifically binds a polypeptide encoded by *ADAMTS2* nucleotide sequence; and an antibody that specifically binds to a polypeptide encoded by *ADAMTS2* nucleotide sequence or binds to a nucleic acid having such a nucleotide sequence. In specific embodiments, the osteoarthritis directed molecule interacts with a nucleic acid or polypeptide variant associated with osteoarthritis, such as variants referenced herein. In other embodiments, the osteoarthritis directed molecule interacts with a polypeptide involved in a signal pathway of a polypeptide encoded by a *ADAMTS2* nucleotide sequence, or a nucleic acid comprising such a nucleotide sequence.

[0109] Compositions sometimes include an adjuvant known to stimulate an immune response, and in certain embodiments, an adjuvant that stimulates a T-cell lymphocyte response. Adjuvants are known, including but not limited to an aluminum adjuvant (e.g., aluminum hydroxide); a cytokine adjuvant or adjuvant that stimulates a cytokine response (e.g., interleukin (IL)-12 and/or gamma-interferon cytokines); a Freund-type mineral oil adjuvant emulsion (e.g., Freund's complete or incomplete adjuvant); a synthetic lipid compound; a copolymer adjuvant (e.g., TitreMax); a saponin; Quil A; a liposome; an oil-in-water emulsion (e.g., an emulsion stabilized by Tween 80 and pluronic polyoxyethylene/polyoxypropylene block copolymer (Syntex Adjuvant Formulation); TitreMax; detoxified endotoxin (MPL) and mycobacterial cell wall components (TDW, CWS) in 2% squalene (Ribi Adjuvant System)); a muramyl dipeptide; an immune-stimulating complex (ISCOM, e.g., an Ag-modified saponin/cholesterol micelle that forms stable cage-like structure); an aqueous phase adjuvant that does not have a depot effect (e.g., Gerbu adjuvant); a carbohydrate polymer (e.g., AdjuPrime); L-tyrosine; a manide-oleate compound (e.g., Montanide); an ethylene-vinyl acetate copolymer (e.g., Elvax 40W1,2); or lipid A, for example. Such compositions are useful for generating an immune response against osteoarthritis directed molecule (e.g., an HLA-binding subsequence within a polypeptide encoded by a *ADAMTS2* nucleotide sequence). In such methods, a peptide having an amino acid subsequence of a polypeptide encoded by a *ADAMTS2* nucleotide sequence is delivered to a subject, where the subsequence binds to an HLA molecule and induces a CTL lymphocyte response. The peptide sometimes is delivered to the subject as an isolated peptide or as a minigene in a plasmid that encodes the

peptide. Methods for identifying HLA-binding subsequences in such polypeptides are known (see e.g., publication WO02/20616 and PCT application US98/01373 for methods of identifying such sequences).

[0110] The cell may be in a group of cells cultured *in vitro* or in a tissue maintained *in vitro* or present in an animal *in vivo* (e.g., a rat, mouse, ape or human). In certain embodiments, a composition comprises a component from a cell such as a nucleic acid molecule (e.g., genomic DNA), a protein mixture or isolated protein, for example. The aforementioned compositions have utility in diagnostic, prognostic and pharmacogenomic methods described previously and in therapeutics described hereafter. Certain osteoarthritis directed molecules are described in greater detail below.

### Compounds

[0111] Compounds can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; peptoid libraries (libraries of molecules having the functionalities of peptides, but with a novel, non-peptide backbone which are resistant to enzymatic degradation but which nevertheless remain bioactive (see, e.g., Zuckermann et al., J. Med. Chem. 37: 2678-85 (1994)); spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; "one-bead one-compound" library methods; and synthetic library methods using affinity chromatography selection. Biological library and peptoid library approaches are typically limited to peptide libraries, while the other approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam, Anticancer Drug Des. 12: 145, (1997)). Examples of methods for synthesizing molecular libraries are described, for example, in DeWitt et al., Proc. Natl. Acad. Sci. U.S.A. 90: 6909 (1993); Erb et al., Proc. Natl. Acad. Sci. USA 91: 11422 (1994); Zuckermann et al., J. Med. Chem. 37: 2678 (1994); Cho et al., Science 261: 1303 (1993); Carrell et al., Angew. Chem. Int. Ed. Engl. 33: 2059 (1994); Carrell et al., Angew. Chem. Int. Ed. Engl. 33: 2061 (1994); and in Gallop et al., J. Med. Chem. 37: 1233 (1994).

[0112] Libraries of compounds may be presented in solution (e.g., Houghten, Biotechniques 13: 412-421 (1992)), or on beads (Lam, Nature 354: 82-84 (1991)), chips (Fodor, Nature 364: 555-556 (1993)), bacteria or spores (Ladner, United States Patent No. 5,223,409), plasmids (Cull et al., Proc. Natl. Acad. Sci. USA 89: 1865-1869 (1992)) or on phage (Scott and Smith, Science 249: 386-390 (1990); Devlin, Science 249: 404-406 (1990); Cwirla et al., Proc. Natl. Acad. Sci. 87: 6378-6382 (1990); Felici, J. Mol. Biol. 222: 301-310 (1991); Ladner supra.).

[0113] A compound sometimes alters expression and sometimes alters activity of a polypeptide target and may be a small molecule. Small molecules include, but are not limited to, peptides, peptidomimetics (e.g., peptoids), amino acids, amino acid analogs, polynucleotides, polynucleotide analogs, nucleotides, nucleotide analogs, organic or inorganic compounds (i.e., including heteroorganic and organometallic compounds) having a molecular weight less than about 10,000 grams per mole,

organic or inorganic compounds having a molecular weight less than about 5,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 1,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 500 grams per mole, and salts, esters, and other pharmaceutically acceptable forms of such compounds.

Antisense Nucleic Acid Molecules, Ribozymes, RNAi, siRNA and Modified Nucleic Acid Molecules

[0114] An “antisense” nucleic acid refers to a nucleotide sequence complementary to a “sense” nucleic acid encoding a polypeptide, e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. The antisense nucleic acid can be complementary to an entire coding strand, or to a portion thereof or a substantially identical sequence thereof. In another embodiment, the antisense nucleic acid molecule is antisense to a “noncoding region” of the coding strand of a nucleotide sequence (e.g., 5’ and 3’ untranslated regions in SEQ ID NO: 1).

[0115] An antisense nucleic acid can be designed such that it is complementary to the entire coding region of an mRNA encoded by a nucleotide sequence (e.g., SEQ ID NO: 1), and often the antisense nucleic acid is an oligonucleotide antisense to only a portion of a coding or noncoding region of the mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of the mRNA, e.g., between the -10 and +10 regions of the target gene nucleotide sequence of interest. An antisense oligonucleotide can be, for example, about 7, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, or more nucleotides in length. The antisense nucleic acids, which include the ribozymes described hereafter, can be designed to target a *ADAMTS2* nucleotide sequence, often a variant associated with osteoarthritis, or a substantially identical sequence thereof. Among the variants, minor alleles and major alleles can be targeted, and those associated with a higher risk of osteoarthritis are often designed, tested, and administered to subjects.

[0116] An antisense nucleic acid can be constructed using chemical synthesis and enzymatic ligation reactions using standard procedures. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. Antisense nucleic acid also can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

[0117] When utilized as therapeutics, antisense nucleic acids typically are administered to a subject (e.g., by direct injection at a tissue site) or generated in situ such that they hybridize with or bind to

cellular mRNA and/or genomic DNA encoding a polypeptide and thereby inhibit expression of the polypeptide, for example, by inhibiting transcription and/or translation. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then are administered systemically. For systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, for example, by linking antisense nucleic acid molecules to peptides or antibodies which bind to cell surface receptors or antigens. Antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. Sufficient intracellular concentrations of antisense molecules are achieved by incorporating a strong promoter, such as a pol II or pol III promoter, in the vector construct.

[0118] Antisense nucleic acid molecules sometimes are alpha-anomeric nucleic acid molecules. An alpha-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual beta-units, the strands run parallel to each other (Gaultier et al., *Nucleic Acids. Res.* 15: 6625-6641 (1987)). Antisense nucleic acid molecules can also comprise a 2'-o-methylribonucleotide (Inoue et al., *Nucleic Acids Res.* 15: 6131-6148 (1987)) or a chimeric RNA-DNA analogue (Inoue et al., *FEBS Lett.* 215: 327-330 (1987)). Antisense nucleic acids sometimes are composed of DNA or PNA or any other nucleic acid derivatives described previously.

[0119] In another embodiment, an antisense nucleic acid is a ribozyme. A ribozyme having specificity for a *ADAMTS2* nucleotide sequence can include one or more sequences complementary to such a nucleotide sequence, and a sequence having a known catalytic region responsible for mRNA cleavage (see e.g., U.S. Pat. No. 5,093,246 or Haselhoff and Gerlach, *Nature* 334: 585-591 (1988)). For example, a derivative of a *Tetrahymena* L-19 IVS RNA is sometimes utilized in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a mRNA (see e.g., Cech et al. U.S. Patent No. 4,987,071; and Cech et al. U.S. Patent No. 5,116,742). Also, target mRNA sequences can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules (see e.g., Bartel & Szostak, *Science* 261: 1411-1418 (1993)).

[0120] Osteoarthritis directed molecules include in certain embodiments nucleic acids that can form triple helix structures with a *ADAMTS2* nucleotide sequence, or a substantially identical sequence thereof, especially one that includes a regulatory region that controls expression of a polypeptide. Gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of a nucleotide sequence referenced herein or a substantially identical sequence (e.g., promoter and/or enhancers) to form triple helical structures that prevent transcription of a gene in target cells (see e.g., Helene, *Anticancer Drug Des.* 6(6): 569-84 (1991); Helene et al., *Ann. N.Y. Acad. Sci.* 660: 27-36 (1992); and Maher, *Bioassays* 14(12): 807-15 (1992). Potential sequences that can be targeted for triple helix formation can be increased by creating a so-called "switchback" nucleic acid molecule. Switchback molecules are synthesized in an alternating 5'-3', 3'-5' manner, such that they base pair with first one

strand of a duplex and then the other, eliminating the necessity for a sizeable stretch of either purines or pyrimidines to be present on one strand of a duplex.

[0121] Osteoarthritis directed molecules include RNAi and siRNA nucleic acids. Gene expression may be inhibited by the introduction of double-stranded RNA (dsRNA), which induces potent and specific gene silencing, a phenomenon called RNA interference or RNAi. See, e.g., Fire et al., US Patent Number 6,506,559; Tuschl et al. PCT International Publication No. WO 01/75164; Kay et al. PCT International Publication No. WO 03/010180A1; or Bosher JM, Labouesse, Nat Cell Biol 2000 Feb;2(2):E31-6. This process has been improved by decreasing the size of the double-stranded RNA to 20-24 base pairs (to create small-interfering RNAs or siRNAs) that “switched off” genes in mammalian cells without initiating an acute phase response, i.e., a host defense mechanism that often results in cell death (see, e.g., Caplen et al. Proc Natl Acad Sci U S A. 2001 Aug 14;98(17):9742-7 and Elbashir et al. Methods 2002 Feb;26(2):199-213). There is increasing evidence of post-transcriptional gene silencing by RNA interference (RNAi) for inhibiting targeted expression in mammalian cells at the mRNA level, in human cells. There is additional evidence of effective methods for inhibiting the proliferation and migration of tumor cells in human patients, and for inhibiting metastatic cancer development (see, e.g., U.S. Patent Application No. US2001000993183; Caplen et al. Proc Natl Acad Sci U S A; and Abderrahmani et al. Mol Cell Biol 2001 Nov21(21):7256-67).

[0122] An “siRNA” or “RNAi” refers to a nucleic acid that forms a double stranded RNA and has the ability to reduce or inhibit expression of a gene or target gene when the siRNA is delivered to or expressed in the same cell as the gene or target gene. “siRNA” refers to short double-stranded RNA formed by the complementary strands. Complementary portions of the siRNA that hybridize to form the double stranded molecule often have substantial or complete identity to the target molecule sequence. In one embodiment, an siRNA refers to a nucleic acid that has substantial or complete identity to a target gene and forms a double stranded siRNA.

[0123] When designing the siRNA molecules, the targeted region often is selected from a given DNA sequence beginning 50 to 100 nucleotides downstream of the start codon. See, e.g., Elbashir et al., Methods 26:199-213 (2002). Initially, 5' or 3' UTRs and regions nearby the start codon were avoided assuming that UTR-binding proteins and/or translation initiation complexes may interfere with binding of the siRNP or RISC endonuclease complex. Sometimes regions of the target 23 nucleotides in length conforming to the sequence motif AA(N19)TT (N, an nucleotide), and regions with approximately 30% to 70% G/C-content (often about 50% G/C-content) often are selected. If no suitable sequences are found, the search often is extended using the motif NA(N21). The sequence of the sense siRNA sometimes corresponds to (N19) TT or N21 (position 3 to 23 of the 23-nt motif), respectively. In the latter case, the 3' end of the sense siRNA often is converted to TT. The rationale for this sequence conversion is to generate a symmetric duplex with respect to the sequence composition of the sense and



antisense 3' overhangs. The antisense siRNA is synthesized as the complement to position 1 to 21 of the 23-nt motif. Because position 1 of the 23-nt motif is not recognized sequence-specifically by the antisense siRNA, the 3'-most nucleotide residue of the antisense siRNA can be chosen deliberately. However, the penultimate nucleotide of the antisense siRNA (complementary to position 2 of the 23-nt motif) often is complementary to the targeted sequence. For simplifying chemical synthesis, TT often is utilized. siRNAs corresponding to the target motif NAR(N17)YNN, where R is purine (A,G) and Y is pyrimidine (C,U), often are selected. Respective 21 nucleotide sense and antisense siRNAs often begin with a purine nucleotide and can also be expressed from pol III expression vectors without a change in targeting site. Expression of RNAs from pol III promoters often is efficient when the first transcribed nucleotide is a purine.

[0124] The sequence of the siRNA can correspond to the full length target gene, or a subsequence thereof. Often, the siRNA is about 15 to about 50 nucleotides in length (e.g., each complementary sequence of the double stranded siRNA is 15-50 nucleotides in length, and the double stranded siRNA is about 15-50 base pairs in length, sometimes about 20-30 nucleotides in length or about 20-25 nucleotides in length, e.g., 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 nucleotides in length. The siRNA sometimes is about 21 nucleotides in length. Methods of using siRNA are well known in the art, and specific siRNA molecules may be purchased from a number of companies including Dharmacon Research, Inc.

[0125] Antisense, ribozyme, RNAi and siRNA nucleic acids can be altered to form modified nucleic acid molecules. The nucleic acids can be altered at base moieties, sugar moieties or phosphate backbone moieties to improve stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of nucleic acid molecules can be modified to generate peptide nucleic acids (see Hyrup et al., *Bioorganic & Medicinal Chemistry* 4 (1): 5-23 (1996)). As used herein, the terms "peptide nucleic acid" or "PNA" refers to a nucleic acid mimic such as a DNA mimic, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of a PNA can allow for specific hybridization to DNA and RNA under conditions of low ionic strength. Synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described, for example, in Hyrup et al., (1996) *supra* and Perry-O'Keefe et al., *Proc. Natl. Acad. Sci.* 93: 14670-675 (1996).

[0126] PNA nucleic acids can be used in prognostic, diagnostic, and therapeutic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, for example, inducing transcription or translation arrest or inhibiting replication. PNA nucleic acid molecules can also be used in the analysis of single base pair mutations in a gene, (e.g., by PNA-directed PCR clamping); as "artificial restriction enzymes" when used in combination with other enzymes, (e.g., S1 nucleases (Hyrup (1996) *supra*)); or as probes or primers for DNA sequencing or hybridization (Hyrup et al., (1996) *supra*; Perry-O'Keefe *supra*).

[0127] In other embodiments, oligonucleotides may include other appended groups such as peptides (e.g., for targeting host cell receptors in vivo), or agents facilitating transport across cell membranes (see e.g., Letsinger et al., Proc. Natl. Acad. Sci. USA 86: 6553-6556 (1989); Lemaitre et al., Proc. Natl. Acad. Sci. USA 84: 648-652 (1987); PCT Publication No. W088/09810) or the blood-brain barrier (see, e.g., PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (See, e.g., Krol et al., Bio-Techniques 6: 958-976 (1988)) or intercalating agents. (See, e.g., Zon, Pharm. Res. 5: 539-549 (1988) ). To this end, the oligonucleotide may be conjugated to another molecule, (e.g., a peptide, hybridization triggered cross-linking agent, transport agent, or hybridization-triggered cleavage agent).

[0128] Also included herein are molecular beacon oligonucleotide primer and probe molecules having one or more regions complementary to a *ADAMTS2* nucleotide sequence, or a substantially identical sequence thereof, two complementary regions one having a fluorophore and one a quencher such that the molecular beacon is useful for quantifying the presence of the nucleic acid in a sample. Molecular beacon nucleic acids are described, for example, in Lizardi et al., U.S. Patent No. 5,854,033; Nazarenko et al., U.S. Patent No. 5,866,336, and Livak et al., U.S. Patent 5,876,930.

#### Antibodies

[0129] The term “antibody” as used herein refers to an immunoglobulin molecule or immunologically active portion thereof, i.e., an antigen-binding portion. Examples of immunologically active portions of immunoglobulin molecules include F(ab) and F(ab')<sub>2</sub> fragments which can be generated by treating the antibody with an enzyme such as pepsin. An antibody sometimes is a polyclonal, monoclonal, recombinant (e.g., a chimeric or humanized), fully human, non-human (e.g., murine), or a single chain antibody. An antibody may have effector function and can fix complement, and is sometimes coupled to a toxin or imaging agent.

[0130] A full-length polypeptide or antigenic peptide fragment encoded by a nucleotide sequence referenced herein can be used as an immunogen or can be used to identify antibodies made with other immunogens, e.g., cells, membrane preparations, and the like. An antigenic peptide often includes at least 8 amino acid residues of the amino acid sequences encoded by a nucleotide sequence referenced herein, or substantially identical sequence thereof, and encompasses an epitope. Antigenic peptides sometimes include 10 or more amino acids, 15 or more amino acids, 20 or more amino acids, or 30 or more amino acids. Hydrophilic and hydrophobic fragments of polypeptides sometimes are used as immunogens.

[0131] Epitopes encompassed by the antigenic peptide are regions located on the surface of the polypeptide (e.g., hydrophilic regions) as well as regions with high antigenicity. For example, an Emini surface probability analysis of the human polypeptide sequence can be used to indicate the regions that

have a particularly high probability of being localized to the surface of the polypeptide and are thus likely to constitute surface residues useful for targeting antibody production. The antibody may bind an epitope on any domain or region on polypeptides described herein.

[0132] Also, chimeric, humanized, and completely human antibodies are useful for applications which include repeated administration to subjects. Chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, can be made using standard recombinant DNA techniques. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in Robinson et al International Application No. PCT/US86/02269; Akira, et al European Patent Application 184,187; Taniguchi, M., European Patent Application 171,496; Morrison et al European Patent Application 173,494; Neuberger et al PCT International Publication No. WO 86/01533; Cabilly et al U.S. Patent No. 4,816,567; Cabilly et al European Patent Application 125,023; Better et al., Science 240: 1041-1043 (1988); Liu et al., Proc. Natl. Acad. Sci. USA 84: 3439-3443 (1987); Liu et al., J. Immunol. 139: 3521-3526 (1987); Sun et al., Proc. Natl. Acad. Sci. USA 84: 214-218 (1987); Nishimura et al., Canc. Res. 47: 999-1005 (1987); Wood et al., Nature 314: 446-449 (1985); and Shaw et al., J. Natl. Cancer Inst. 80: 1553-1559 (1988); Morrison, S. L., Science 229: 1202-1207 (1985); Oi et al., BioTechniques 4: 214 (1986); Winter U.S. Patent 5,225,539; Jones et al., Nature 321: 552-525 (1986); Verhoeyan et al., Science 239: 1534; and Beidler et al., J. Immunol. 141: 4053-4060 (1988).

[0133] Completely human antibodies are particularly desirable for therapeutic treatment of human patients. Such antibodies can be produced using transgenic mice that are incapable of expressing endogenous immunoglobulin heavy and light chains genes, but which can express human heavy and light chain genes. See, for example, Lonberg and Huszar, Int. Rev. Immunol. 13: 65-93 (1995); and U.S. Patent Nos. 5,625,126; 5,633,425; 5,569,825; 5,661,016; and 5,545,806. In addition, companies such as Abgenix, Inc. (Fremont, CA) and Medarex, Inc. (Princeton, NJ), can be engaged to provide human antibodies directed against a selected antigen using technology similar to that described above. Completely human antibodies that recognize a selected epitope also can be generated using a technique referred to as "guided selection." In this approach a selected non-human monoclonal antibody (e.g., a murine antibody) is used to guide the selection of a completely human antibody recognizing the same epitope. This technology is described for example by Jespers et al., Bio/Technology 12: 899-903 (1994).

[0134] An antibody can be a single chain antibody. A single chain antibody (scFV) can be engineered (see, e.g., Colcher et al., Ann. N Y Acad. Sci. 880: 263-80 (1999); and Reiter, Clin. Cancer Res. 2: 245-52 (1996)). Single chain antibodies can be dimerized or multimerized to generate multivalent antibodies having specificities for different epitopes of the same target polypeptide.

[0135] Antibodies also may be selected or modified so that they exhibit reduced or no ability to bind an Fc receptor. For example, an antibody may be an isotype or subtype, fragment or other mutant, which

does not support binding to an Fc receptor (e.g., it has a mutagenized or deleted Fc receptor binding region).

[0136] Also, an antibody (or fragment thereof) may be conjugated to a therapeutic moiety such as a cytotoxin, a therapeutic agent or a radioactive metal ion. A cytotoxin or cytotoxic agent includes any agent that is detrimental to cells. Examples include taxol, cytochalasin B, gramicidin D, ethidium bromide, emetine, mitomycin, etoposide, teniposide, vincristine, vinblastine, colchicin, doxorubicin, daunorubicin, dihydroxy anthracin dione, mitoxantrone, mithramycin, actinomycin D, 1 dehydrotestosterone, glucocorticoids, procaine, tetracaine, lidocaine, propranolol, and puromycin and analogs or homologs thereof. Therapeutic agents include, but are not limited to, antimetabolites (e.g., methotrexate, 6-mercaptopurine, 6-thioguanine, cytarabine, 5-fluorouracil decarbazine), alkylating agents (e.g., mechlorethamine, thiotepe chlorambucil, melphalan, carmustine (BCNU) and lomustine (CCNU), cyclophosphamide, busulfan, dibromomannitol, streptozotocin, mitomycin C, and cis-dichlorodiamine platinum (II) (DDP) cisplatin), anthracyclines (e.g., daunorubicin (formerly daunomycin) and doxorubicin), antibiotics (e.g., dactinomycin (formerly actinomycin), bleomycin, mithramycin, and anthramycin (AMC)), and anti-mitotic agents (e.g., vincristine and vinblastine).

[0137] Antibody conjugates can be used for modifying a given biological response. For example, the drug moiety may be a protein or polypeptide possessing a desired biological activity. Such proteins may include, for example, a toxin such as abrin, ricin A, pseudomonas exotoxin, or diphtheria toxin; a polypeptide such as tumor necrosis factor, gamma-interferon, alpha-interferon, nerve growth factor, platelet derived growth factor, tissue plasminogen activator; or, biological response modifiers such as, for example, lymphokines, interleukin-1 ("IL-1"), interleukin-2 ("IL-2"), interleukin-6 ("IL-6"), granulocyte macrophage colony stimulating factor ("GM-CSF"), granulocyte colony stimulating factor ("G-CSF"), or other growth factors. Also, an antibody can be conjugated to a second antibody to form an antibody heteroconjugate as described by Segal in U.S. Patent No. 4,676,980, for example.

[0138] An antibody (e.g., monoclonal antibody) can be used to isolate target polypeptides by standard techniques, such as affinity chromatography or immunoprecipitation. Moreover, an antibody can be used to detect a target polypeptide (e.g., in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the polypeptide. Antibodies can be used diagnostically to monitor polypeptide levels in tissue as part of a clinical testing procedure, e.g., to determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance (i.e., antibody labeling). Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase,  $\beta$ -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials

include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include  $^{125}\text{I}$ ,  $^{131}\text{I}$ ,  $^{35}\text{S}$  or  $^3\text{H}$ . Also, an antibody can be utilized as a test molecule for determining whether it can treat osteoarthritis, and as a therapeutic for administration to a subject for treating osteoarthritis.

[0139] An antibody can be made by immunizing with a purified antigen, or a fragment thereof, e.g., a fragment described herein, a membrane associated antigen, tissues, e.g., crude tissue preparations, whole cells, preferably living cells, lysed cells, or cell fractions.

[0140] Included herein are antibodies which bind only a native polypeptide, only denatured or otherwise non-native polypeptide, or which bind both, as well as those having linear or conformational epitopes. Conformational epitopes sometimes can be identified by selecting antibodies that bind to native but not denatured polypeptide. Also featured are antibodies that specifically bind to a polypeptide variant associated with osteoarthritis.

#### Methods for Identifying Candidate Therapeutics for Treating Osteoarthritis

[0141] Current therapies for the treatment of osteoarthritis have limited efficacy, limited tolerability and significant mechanism-based side effects, and few of the available therapies adequately address underlying defects. Current therapeutic approaches were largely developed in the absence of defined molecular targets or even a solid understanding of disease pathogenesis. Therefore, provided are methods of identifying candidate therapeutics that target biochemical pathways related to the development of osteoarthritis.

[0142] Thus, featured herein are methods for identifying a candidate therapeutic for treating osteoarthritis. The methods comprise contacting a test molecule with a target molecule in a system. A “target molecule” as used herein refers to a *ADAMTS2* nucleic acid, a substantially identical nucleic acid thereof, or a fragment thereof, and an encoded polypeptide of the foregoing. The methods also comprise determining the presence or absence of an interaction between the test molecule and the target molecule, where the presence of an interaction between the test molecule and the nucleic acid or polypeptide identifies the test molecule as a candidate osteoarthritis therapeutic. The interaction between the test molecule and the target molecule may be quantified.

[0143] Test molecules and candidate therapeutics include, but are not limited to, compounds, antisense nucleic acids, siRNA molecules, ribozymes, polypeptides or proteins encoded by a *ADAMTS2* nucleotide sequence, or a substantially identical sequence or fragment thereof, and immunotherapeutics (e.g., antibodies and HLA-presented polypeptide fragments). A test molecule or candidate therapeutic may act as a modulator of target molecule concentration or target molecule function in a system. A

“modulator” may agonize (i.e., up-regulates) or antagonize (i.e., down-regulates) a target molecule concentration partially or completely in a system by affecting such cellular functions as DNA replication and/or DNA processing (e.g., DNA methylation or DNA repair), RNA transcription and/or RNA processing (e.g., removal of intronic sequences and/or translocation of spliced mRNA from the nucleus), polypeptide production (e.g., translation of the polypeptide from mRNA), and/or polypeptide post-translational modification (e.g., glycosylation, phosphorylation, and proteolysis of pro-polypeptides). A modulator may also agonize or antagonize a biological function of a target molecule partially or completely, where the function may include adopting a certain structural conformation, interacting with one or more binding partners, ligand binding, catalysis (e.g., phosphorylation, dephosphorylation, hydrolysis, methylation, and isomerization), and an effect upon a cellular event (e.g., effecting progression of osteoarthritis).

[0144] As used herein, the term “system” refers to a cell free *in vitro* environment and a cell-based environment such as a collection of cells, a tissue, an organ, or an organism. A system is “contacted” with a test molecule in a variety of manners, including adding molecules in solution and allowing them to interact with one another by diffusion, cell injection, and any administration routes in an animal. As used herein, the term “interaction” refers to an effect of a test molecule on test molecule, where the effect sometimes is binding between the test molecule and the target molecule, and sometimes is an observable change in cells, tissue, or organism.

[0145] There are many standard methods for detecting the presence or absence of interaction between a test molecule and a target molecule. For example, titrametric, acidimetric, radiometric, NMR, monolayer, polarographic, spectrophotometric, fluorescent, and ESR assays probative of a target molecule interaction may be utilized.

[0146] *ADAMTS2* activity and/or *ADAMTS2* interactions can be detected and quantified using assays known in the art. For example, an immunoprecipitation assay or a kinase activity assay that employs a kinase-inactivated MEK can be utilized. Kinase inactivated MEKs are known in the art, such as a MEK that includes the mutation K97M. In these assays, mammalian cells (e.g., COS or NIH-3T3) are transiently transfected with constructs expressing *ADAMTS2*, and in addition, the cells are co-transfected with oncogenic RAS or SRC or both. Oncogenic RAS or SRC activates *ADAMTS2* kinase activity. *ADAMTS2* is immunoprecipitated from cell extracts using a monoclonal antibody (e.g., 9E10) or a polyclonal antibody (e.g., from rabbit) specific for a unique peptide from *ADAMTS2*. *ADAMTS2* is then resuspended in assay buffer containing GST-Mek1 or GST-Mek2 and/or GST-ERK2. In addition, [ $\gamma$   $^{32}$ P] ATP can be added to detect and/or quantify phosphorylation activity. Samples are incubated for 5-30 minutes at 30°C, and then the reaction is terminated by addition of EDTA. The samples are centrifuged and the supernatant fractions are collected. Phosphorylation activity is detected using one of two methods: (i) activity of GST-ERK2 kinase can be measured using MBP (myelin basic

protein, a substrate for ERK) as substrate, or (ii) following incubation of immunoprecipitated *ADAMTS2* in reaction buffer containing GST-ERK and [ $\gamma$   $^{32}\text{P}$ ] ATP, transfer of labeled ATP to kinase-dead ERK can be quantified by a phosphor-imager or densitometer following PAGE separation of polypeptide products (phosphorylated and non-phosphorylated forms). These types of assays are described in Weber et al., *Oncogene* 19: 169-176 (2000); Mason et al., *EMBO J.* 18: 2137-2148 (1999); Marais et al., *J. Biol. Chem.* 272: 4378-4383 (1997); Marais et al., *EMBO J.* 14: 3136-3145 (1995).

[0147] As noted above, *ADAMTS2* includes a domain having metalloprotease activity, and modulators of such activity are known. Examples of such modulators are set forth in WO03063762A2; WO-09937625; WO-09918076; WO-09838163; WO-09837877; WO9947550A1; WO0177092A1; WO0040577A1; WO9942436A1; WO9838163A1; WO9837877A1; WO04014379A1; WO03106381A2; WO03014098A1; WO03014092A1 and WO02096426A1.

[0148] Test molecule/target molecule interactions can be detected and/or quantified using assays known in the art. For example, an interaction can be determined by labeling the test molecule and/or the target molecule, where the label is covalently or non-covalently attached to the test molecule or target molecule. The label is sometimes a radioactive molecule such as  $^{125}\text{I}$ ,  $^{131}\text{I}$ ,  $^{35}\text{S}$  or  $^3\text{H}$ , which can be detected by direct counting of radioemission or by scintillation counting. Also, enzymatic labels such as horseradish peroxidase, alkaline phosphatase, or luciferase may be utilized where the enzymatic label can be detected by determining conversion of an appropriate substrate to product. In addition, presence or absence of an interaction can be determined without labeling. For example, a microphysiometer (*e.g.*, Cytosensor) is an analytical instrument that measures the rate at which a cell acidifies its environment using a light-addressable potentiometric sensor (LAPS). Changes in this acidification rate can be used as an indication of an interaction between a test molecule and target molecule (McConnell, H. M. *et al.*, *Science* 257: 1906-1912 (1992)).

[0149] In cell-based systems, cells typically include a *ADAMTS2* nucleic acid, an encoded polypeptide, or substantially identical nucleic acid or polypeptide thereof, and are often of mammalian origin, although the cell can be of any origin. Whole cells, cell homogenates, and cell fractions (*e.g.*, cell membrane fractions) can be subjected to analysis. Where interactions between a test molecule with a target polypeptide are monitored, soluble and/or membrane bound forms of the polypeptide may be utilized. Where membrane-bound forms of the polypeptide are used, it may be desirable to utilize a solubilizing agent. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton<sup>®</sup> X-100, Triton<sup>®</sup> X-114, Thesit<sup>®</sup>, Isotridecypoly(ethylene glycol ether)<sub>n</sub>, 3-[(3-cholamidopropyl)dimethylamminio]-1-propane sulfonate (CHAPS), 3-[(3-

cholamidopropyl)dimethylamminio]-2-hydroxy-1-propane sulfonate (CHAPSO), or N-dodecyl-N,N-dimethyl-3-ammonio-1-propane sulfonate.

[0150] An interaction between a test molecule and target molecule also can be detected by monitoring fluorescence energy transfer (FET) (*see, e.g.,* Lakowicz *et al.*, U.S. Patent No. 5,631,169; Stavrianopoulos *et al.* U.S. Patent No. 4,868,103). A fluorophore label on a first, “donor” molecule is selected such that its emitted fluorescent energy will be absorbed by a fluorescent label on a second, “acceptor” molecule, which in turn is able to fluoresce due to the absorbed energy. Alternately, the “donor” polypeptide molecule may simply utilize the natural fluorescent energy of tryptophan residues. Labels are chosen that emit different wavelengths of light, such that the “acceptor” molecule label may be differentiated from that of the “donor”. Since the efficiency of energy transfer between the labels is related to the distance separating the molecules, the spatial relationship between the molecules can be assessed. In a situation in which binding occurs between the molecules, the fluorescent emission of the “acceptor” molecule label in the assay should be maximal. An FET binding event can be conveniently measured through standard fluorometric detection means well known in the art (*e.g.,* using a fluorimeter).

[0151] In another embodiment, determining the presence or absence of an interaction between a test molecule and a target molecule can be effected by monitoring surface plasmon resonance (*see, e.g.,* Sjolander & Urbanicz, *Anal. Chem.* 63: 2338-2345 (1991) and Szabo *et al.*, *Curr. Opin. Struct. Biol.* 5: 699-705 (1995)). “Surface plasmon resonance” or “biomolecular interaction analysis (BIA)” can be utilized to detect biospecific interactions in real time, without labeling any of the interactants (*e.g.,* BIAcore). Changes in the mass at the binding surface (indicative of a binding event) result in alterations of the refractive index of light near the surface (the optical phenomenon of surface plasmon resonance (SPR)), resulting in a detectable signal which can be used as an indication of real-time reactions between biological molecules.

[0152] In another embodiment, the target molecule or test molecules are anchored to a solid phase, facilitating the detection of target molecule/test molecule complexes and separation of the complexes from free, uncomplexed molecules. The target molecule or test molecule is immobilized to the solid support. In an embodiment, the target molecule is anchored to a solid surface, and the test molecule, which is not anchored, can be labeled, either directly or indirectly, with detectable labels discussed herein.

[0153] It may be desirable to immobilize a target molecule, an anti-target molecule antibody, and/or test molecules to facilitate separation of target molecule/test molecule complexes from uncomplexed forms, as well as to accommodate automation of the assay. The attachment between a test molecule and/or target molecule and the solid support may be covalent or non-covalent (*see, e.g.,* U.S. Patent No. 6,022,688 for non-covalent attachments). The solid support may be one or more surfaces of the system, such as one or more surfaces in each well of a microtiter plate, a surface of a silicon wafer, a surface of a



bead (*see, e.g.*, Lam, *Nature* 354: 82-84 (1991)) that is optionally linked to another solid support, or a channel in a microfluidic device, for example. Types of solid supports, linker molecules for covalent and non-covalent attachments to solid supports, and methods for immobilizing nucleic acids and other molecules to solid supports are well known (*see, e.g.*, U.S. Patent Nos. 6,261,776; 5,900,481; 6,133,436; and 6,022,688; and WIPO publication WO 01/18234).

**[0154]** In an embodiment, target molecule may be immobilized to surfaces via biotin and streptavidin. For example, biotinylated target polypeptide can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques known in the art (*e.g.*, biotinylation kit, Pierce Chemicals, Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). In another embodiment, a target polypeptide can be prepared as a fusion polypeptide. For example, glutathione-S-transferase/target polypeptide fusion can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivitized microtiter plates, which are then combined with a test molecule under conditions conducive to complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components, or the matrix is immobilized in the case of beads, and complex formation is determined directly or indirectly as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of target molecule binding or activity is determined using standard techniques.

**[0155]** In an embodiment, the non-immobilized component is added to the coated surface containing the anchored component. After the reaction is complete, unreacted components are removed (*e.g.*, by washing) under conditions such that a significant percentage of complexes formed will remain immobilized to the solid surface. The detection of complexes anchored on the solid surface can be accomplished in a number of manners. Where the previously non-immobilized component is pre-labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the previously non-immobilized component is not pre-labeled, an indirect label can be used to detect complexes anchored on the surface, *e.g.*, by adding a labeled antibody specific for the immobilized component, where the antibody, in turn, can be directly labeled or indirectly labeled with, *e.g.*, a labeled anti-Ig antibody.

**[0156]** In another embodiment, an assay is performed utilizing antibodies that specifically bind target molecule or test molecule but do not interfere with binding of the target molecule to the test molecule. Such antibodies can be derivitized to a solid support, and unbound target molecule may be immobilized by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the target molecule.

[0157] Cell free assays also can be conducted in a liquid phase. In such an assay, reaction products are separated from unreacted components, by any of a number of standard techniques, including but not limited to: differential centrifugation (*see, e.g.,* Rivas, G., and Minton, *Trends Biochem Sci Aug:18(8): 284-7 (1993)*); chromatography (gel filtration chromatography, ion-exchange chromatography); electrophoresis (*see, e.g.,* Ausubel *et al., eds. Current Protocols in Molecular Biology*, J. Wiley: New York (1999)); and immunoprecipitation (*see, e.g.,* Ausubel *et al., eds., supra*). Media and chromatographic techniques are known to one skilled in the art (*see, e.g.,* Heegaard, *J Mol. Recognit. Winter; 11(1-6): 141-8 (1998)*; Hage & Tweed, *J. Chromatogr. B Biomed. Sci. Appl. Oct 10; 699 (1-2): 499-525 (1997)*). Further, fluorescence energy transfer may also be conveniently utilized, as described herein, to detect binding without further purification of the complex from solution.

[0158] In another embodiment, modulators of target molecule expression are identified. For example, a cell or cell free mixture is contacted with a candidate compound and the expression of target mRNA or target polypeptide is evaluated relative to the level of expression of target mRNA or target polypeptide in the absence of the candidate compound. When expression of target mRNA or target polypeptide is greater in the presence of the candidate compound than in its absence, the candidate compound is identified as an agonist of target mRNA or target polypeptide expression. Alternatively, when expression of target mRNA or target polypeptide is less (*e.g.,* less with statistical significance) in the presence of the candidate compound than in its absence, the candidate compound is identified as an antagonist or inhibitor of target mRNA or target polypeptide expression. The level of target mRNA or target polypeptide expression can be determined by methods described herein.

[0159] In another embodiment, binding partners that interact with a target molecule are detected. The target molecules can interact with one or more cellular or extracellular macromolecules, such as polypeptides *in vivo*, and these interacting molecules are referred to herein as "binding partners." Binding partners can agonize or antagonize target molecule biological activity. Also, test molecules that agonize or antagonize interactions between target molecules and binding partners can be useful as therapeutic molecules as they can up-regulate or down-regulated target molecule activity *in vivo* and thereby treat osteoarthritis.

[0160] Binding partners of target molecules can be identified by methods known in the art. For example, binding partners may be identified by lysing cells and analyzing cell lysates by electrophoretic techniques. Alternatively, a two-hybrid assay or three-hybrid assay can be utilized (*see, e.g.,* U.S. Patent No. 5,283,317; Zervos *et al., Cell* 72:223-232 (1993); Madura *et al., J. Biol. Chem.* 268: 12046-12054 (1993); Bartel *et al., Biotechniques* 14: 920-924 (1993); Iwabuchi *et al., Oncogene* 8: 1693-1696 (1993); and Brent WO94/10300). A two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. The assay often utilizes two different DNA constructs. In one construct, a *ADAMTS2* nucleic acid (sometimes referred to as the

“bait”) is fused to a gene encoding the DNA binding domain of a known transcription factor (*e.g.*, GAL-4). In another construct, a DNA sequence from a library of DNA sequences that encodes a potential binding partner (sometimes referred to as the “prey”) is fused to a gene that encodes an activation domain of the known transcription factor. Sometimes, a *ADAMTS2* nucleic acid can be fused to the activation domain. If the “bait” and the “prey” molecules interact *in vivo*, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (*e.g.*, LacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to identify the potential binding partner.

[0161] In an embodiment for identifying test molecules that antagonize or agonize complex formation between target molecules and binding partners, a reaction mixture containing the target molecule and the binding partner is prepared, under conditions and for a time sufficient to allow complex formation. The reaction mixture often is provided in the presence or absence of the test molecule. The test molecule can be included initially in the reaction mixture, or can be added at a time subsequent to the addition of the target molecule and its binding partner. Control reaction mixtures are incubated without the test molecule or with a placebo. Formation of any complexes between the target molecule and the binding partner then is detected. Decreased formation of a complex in the reaction mixture containing test molecule as compared to in a control reaction mixture indicates that the molecule antagonizes target molecule/binding partner complex formation. Alternatively, increased formation of a complex in the reaction mixture containing test molecule as compared to in a control reaction mixture indicates that the molecule agonizes target molecule/binding partner complex formation. In another embodiment, complex formation of target molecule/binding partner can be compared to complex formation of mutant target molecule/binding partner (*e.g.*, amino acid modifications in a target polypeptide). Such a comparison can be important in those cases where it is desirable to identify test molecules that modulate interactions of mutant but not non-mutated target gene products.

[0162] The assays can be conducted in a heterogeneous or homogeneous format. In heterogeneous assays, target molecule and/or the binding partner are immobilized to a solid phase, and complexes are detected on the solid phase at the end of the reaction. In homogeneous assays, the entire reaction is carried out in a liquid phase. In either approach, the order of addition of reactants can be varied to obtain different information about the molecules being tested. For example, test compounds that agonize target molecule/binding partner interactions can be identified by conducting the reaction in the presence of the test molecule in a competition format. Alternatively, test molecules that agonize preformed complexes, *e.g.*, molecules with higher binding constants that displace one of the components from the complex, can be tested by adding the test compound to the reaction mixture after complexes have been formed.

**[0163]** In a heterogeneous assay embodiment, the target molecule or the binding partner is anchored onto a solid surface (*e.g.*, a microtiter plate), while the non-anchored species is labeled, either directly or indirectly. The anchored molecule can be immobilized by non-covalent or covalent attachments. Alternatively, an immobilized antibody specific for the molecule to be anchored can be used to anchor the molecule to the solid surface. The partner of the immobilized species is exposed to the coated surface with or without the test molecule. After the reaction is complete, unreacted components are removed (*e.g.*, by washing) such that a significant portion of any complexes formed will remain immobilized on the solid surface. Where the non-immobilized species is pre-labeled, the detection of label immobilized on the surface is indicative of complex. Where the non-immobilized species is not pre-labeled, an indirect label can be used to detect complexes anchored to the surface; *e.g.*, by using a labeled antibody specific for the initially non-immobilized species. Depending upon the order of addition of reaction components, test compounds that inhibit complex formation or that disrupt preformed complexes can be detected.

**[0164]** In another embodiment, the reaction can be conducted in a liquid phase in the presence or absence of test molecule, where the reaction products are separated from unreacted components, and the complexes are detected (*e.g.*, using an immobilized antibody specific for one of the binding components to anchor any complexes formed in solution, and a labeled antibody specific for the other partner to detect anchored complexes). Again, depending upon the order of addition of reactants to the liquid phase, test compounds that inhibit complex or that disrupt preformed complexes can be identified.

**[0165]** In an alternate embodiment, a homogeneous assay can be utilized. For example, a preformed complex of the target gene product and the interactive cellular or extracellular binding partner product is prepared. One or both of the target molecule or binding partner is labeled, and the signal generated by the label(s) is quenched upon complex formation (*e.g.*, U.S. Patent No. 4,109,496 that utilizes this approach for immunoassays). Addition of a test molecule that competes with and displaces one of the species from the preformed complex will result in the generation of a signal above background. In this way, test substances that disrupt target molecule/binding partner complexes can be identified.

**[0166]** Candidate therapeutics for treating osteoarthritis are identified from a group of test molecules that interact with a target molecule. Test molecules are normally ranked according to the degree with which they modulate (*e.g.*, agonize or antagonize) a function associated with the target molecule (*e.g.*, DNA replication and/or processing, RNA transcription and/or processing, polypeptide production and/or processing, and/or biological function/activity), and then top ranking modulators are selected. Also, pharmacogenomic information described herein can determine the rank of a modulator. The top 10% of ranked test molecules often are selected for further testing as candidate therapeutics, and sometimes the top 15%, 20%, or 25% of ranked test molecules are selected for further testing as candidate therapeutics. Candidate therapeutics typically are formulated for administration to a subject.

### Therapeutic Formulations

[0167] Formulations and pharmaceutical compositions typically include in combination with a pharmaceutically acceptable carrier one or more target molecule modulators. The modulator often is a test molecule identified as having an interaction with a target molecule by a screening method described above. The modulator may be a compound, an antisense nucleic acid, a ribozyme, an antibody, or a binding partner. Also, formulations may comprise a target polypeptide or fragment thereof in combination with a pharmaceutically acceptable carrier.

[0168] Formulations or pharmaceutical compositions typically include in combination with a pharmaceutically acceptable carrier, a compound, an antisense nucleic acid, a ribozyme, an antibody, a binding partner that interacts with an *ADAMTS2* polypeptide, a *ADAMTS2* nucleic acid, or a fragment thereof. The formulated molecule may be one that is identified by a screening method described above. Also, formulations may comprise a *ADAMTS2* polypeptide or fragment thereof, where the *ADAMTS2* polypeptide contains an isoleucine at position 245 of SEQ ID NO: 4, and a pharmaceutically acceptable carrier. Also, formulations may comprise an active *ADAMTS2* polypeptide or fragment thereof, where *ADAMTS2* polypeptide fragments having activity are selected from amino acids 252-1211, 253-1211, 254-1211, 255-1211, 256-1211, 257-1211, 258-1211, 259-1211 or 260-1211 of SEQ ID NO: 4, where it is understood that the active form of *ADAMTS2* does not contain the propeptide domain. As used herein, the term "pharmaceutically acceptable carrier" includes solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Supplementary active compounds can also be incorporated into the compositions.

[0169] As used herein, the term "pharmaceutically acceptable carrier" includes solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Supplementary active compounds can also be incorporated into the compositions. Pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

[0170] A pharmaceutical composition typically is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*, inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerin, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with

acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

[0171] Oral compositions generally include an inert diluent or an edible carrier. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules, *e.g.*, gelatin capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

[0172] Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It should be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as mannitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

[0173] Sterile injectable solutions can be prepared by incorporating the active compound in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which

yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

[0174] For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, *e.g.*, a gas such as carbon dioxide, or a nebulizer.

[0175] Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art. Molecules can also be prepared in the form of suppositories (*e.g.*, with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

[0176] In one embodiment, active molecules are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. Materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

[0177] It is advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier.

[0178] Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, for determining the LD<sub>50</sub> (the dose lethal to 50% of the population) and the ED<sub>50</sub> (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD<sub>50</sub>/ED<sub>50</sub>. Molecules which exhibit high therapeutic indices are preferred. While molecules that exhibit toxic side effects may be used, care should be taken to design a delivery system

that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

[0179] The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such molecules lies preferably within a range of circulating concentrations that include the ED<sub>50</sub> with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For any molecules used in the methods described herein, the therapeutically effective dose can be estimated initially from cell culture assays. A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the IC<sub>50</sub> (*i.e.*, the concentration of the test compound which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography.

[0180] As defined herein, a therapeutically effective amount of protein or polypeptide (*i.e.*, an effective dosage) ranges from about 0.001 to 30 mg/kg body weight, sometimes about 0.01 to 25 mg/kg body weight, often about 0.1 to 20 mg/kg body weight, and more often about 1 to 10 mg/kg, 2 to 9 mg/kg, 3 to 8 mg/kg, 4 to 7 mg/kg, or 5 to 6 mg/kg body weight. The protein or polypeptide can be administered one time per week for between about 1 to 10 weeks, sometimes between 2 to 8 weeks, often between about 3 to 7 weeks, and more often for about 4, 5, or 6 weeks. The skilled artisan will appreciate that certain factors may influence the dosage and timing required to effectively treat a subject, including but not limited to the severity of the disease or disorder, previous treatments, the general health and/or age of the subject, and other diseases present. Moreover, treatment of a subject with a therapeutically effective amount of a protein, polypeptide, or antibody can include a single treatment or, preferably, can include a series of treatments.

[0181] With regard to polypeptide formulations, featured herein is a method for treating osteoarthritis in a subject, which comprises contacting one or more cells in the subject with a first polypeptide, where the subject comprises a second polypeptide having one or more polymorphic variations associated with cancer, and where the first polypeptide comprises fewer polymorphic variations associated with cancer than the second polypeptide. The first and second polypeptides are encoded by a nucleic acid which comprises a nucleotide sequence in SEQ ID NO: 1-3; a nucleotide sequence which encodes a polypeptide consisting of an amino acid sequence encoded by a nucleotide sequence referenced in SEQ ID NO: 1-3; a nucleotide sequence which encodes a polypeptide that is 90% or more identical to an amino acid sequence encoded by a nucleotide sequence of SEQ ID NO: 1-3 and a nucleotide sequence 90% or more identical to a nucleotide sequence in SEQ ID NO: 1-3. The subject often is a human.



[0182] For antibodies, a dosage of 0.1 mg/kg of body weight (generally 10 mg/kg to 20 mg/kg) is often utilized. If the antibody is to act in the brain, a dosage of 50 mg/kg to 100 mg/kg is often appropriate. Generally, partially human antibodies and fully human antibodies have a longer half-life within the human body than other antibodies. Accordingly, lower dosages and less frequent administration is often possible. Modifications such as lipidation can be used to stabilize antibodies and to enhance uptake and tissue penetration (*e.g.*, into the brain). A method for lipidation of antibodies is described by Cruikshank *et al.*, *J. Acquired Immune Deficiency Syndromes and Human Retrovirology* 14:193 (1997).

[0183] Antibody conjugates can be used for modifying a given biological response, the drug moiety is not to be construed as limited to classical chemical therapeutic agents. For example, the drug moiety may be a protein or polypeptide possessing a desired biological activity. Such proteins may include, for example, a toxin such as abrin, ricin A, pseudomonas exotoxin, or diphtheria toxin; a polypeptide such as tumor necrosis factor, alpha-interferon, beta-interferon, nerve growth factor, platelet derived growth factor, tissue plasminogen activator; or, biological response modifiers such as, for example, lymphokines, interleukin-1 ("IL-1"), interleukin-2 ("IL-2"), interleukin-6 ("IL-6"), granulocyte macrophage colony stimulating factor ("GM-CSF"), granulocyte colony stimulating factor ("G-CSF"), or other growth factors. Alternatively, an antibody can be conjugated to a second antibody to form an antibody heteroconjugate as described by Segal in U.S. Patent No. 4,676,980.

[0184] For compounds, exemplary doses include milligram or microgram amounts of the compound per kilogram of subject or sample weight, for example, about 1 microgram per kilogram to about 500 milligrams per kilogram, about 100 micrograms per kilogram to about 5 milligrams per kilogram, or about 1 microgram per kilogram to about 50 micrograms per kilogram. It is understood that appropriate doses of a small molecule depend upon the potency of the small molecule with respect to the expression or activity to be modulated. When one or more of these small molecules is to be administered to an animal (*e.g.*, a human) in order to modulate expression or activity of a polypeptide or nucleic acid described herein, a physician, veterinarian, or researcher may, for example, prescribe a relatively low dose at first, subsequently increasing the dose until an appropriate response is obtained. In addition, it is understood that the specific dose level for any particular animal subject will depend upon a variety of factors including the activity of the specific compound employed, the age, body weight, general health, gender, and diet of the subject, the time of administration, the route of administration, the rate of excretion, any drug combination, and the degree of expression or activity to be modulated.

[0185] With regard to nucleic acid formulations, gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (*see, e.g.*, U.S. Patent 5,328,470) or by stereotactic injection (*see e.g.*, Chen *et al.*, (1994) *Proc. Natl. Acad. Sci. USA* 91:3054-3057). Pharmaceutical preparations of gene therapy vectors can include a gene therapy vector in an acceptable

diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded.

Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells (e.g., retroviral vectors) the pharmaceutical preparation can include one or more cells which produce the gene delivery system. Examples of gene delivery vectors are described herein.

#### Therapeutic Methods

[0186] A therapeutic formulation described above can be administered to a subject in need of a therapeutic for inducing a desired biological response.. Therapeutic formulations can be administered by any of the paths described herein. With regard to both prophylactic and therapeutic methods of treatment, such treatments may be specifically tailored or modified, based on knowledge obtained from pharmacogenomic analyses described herein.

[0187] As used herein, the term “treatment” is defined as the application or administration of a therapeutic formulation to a subject, or application or administration of a therapeutic agent to an isolated tissue or cell line from a subject with the purpose to cure, heal, alleviate, relieve, alter, remedy, ameliorate, improve or affect osteoarthritis, symptoms of osteoarthritis or a predisposition towards osteoarthritis. A therapeutic formulation includes, but is not limited to, small molecules, peptides, antibodies, ribozymes and antisense oligonucleotides. Administration of a therapeutic formulation can occur prior to the manifestation of symptoms characteristic of osteoarthritis, such that osteoarthritis is prevented or delayed in its progression. The appropriate therapeutic composition can be determined based on screening assays described herein.

[0188] As discussed, successful treatment of osteoarthritis can be brought about by techniques that serve to agonize target molecule expression or function, or alternatively, antagonize target molecule expression or function. These techniques include administration of modulators that include, but are not limited to, small organic or inorganic molecules; antibodies (including, for example, polyclonal, monoclonal, humanized, anti-idiotypic, chimeric or single chain antibodies, and Fab, F(ab')<sub>2</sub> and Fab expression library fragments, scFV molecules, and epitope-binding fragments thereof); and peptides, phosphopeptides, or polypeptides.

[0189] Further, antisense and ribozyme molecules that inhibit expression of the target gene can also be used to reduce the level of target gene expression, thus effectively reducing the level of target gene activity. Still further, triple helix molecules can be utilized in reducing the level of target gene activity. Antisense, ribozyme and triple helix molecules are discussed above. It is possible that the use of antisense, ribozyme, and/or triple helix molecules to reduce or inhibit mutant gene expression can also reduce or inhibit the transcription (triple helix) and/or translation (antisense, ribozyme) of mRNA produced by normal target gene alleles, such that the concentration of normal target gene product present can be lower than is necessary for a normal phenotype. In such cases, nucleic acid molecules that encode

and express target gene polypeptides exhibiting normal target gene activity can be introduced into cells via gene therapy method. Alternatively, in instances in that the target gene encodes an extracellular polypeptide, it can be preferable to co-administer normal target gene polypeptide into the cell or tissue in order to maintain the requisite level of cellular or tissue target gene activity.

[0190] Another method by which nucleic acid molecules may be utilized in treating or preventing osteoarthritis is use of aptamer molecules specific for target molecules. Aptamers are nucleic acid molecules having a tertiary structure which permits them to specifically bind to ligands (*see, e.g., Osborne, et al., Curr. Opin. Chem. Biol.* 1(1): 5-9 (1997); and Patel, D. J., *Curr. Opin. Chem. Biol. Jun*;1(1): 32-46 (1997)).

[0191] Yet another method of utilizing nucleic acid molecules for osteoarthritis treatment is gene therapy, which can also be referred to as allele therapy. Provided herein is a gene therapy method for treating osteoarthritis in a subject, which comprises contacting one or more cells in the subject or from the subject with a nucleic acid having a first nucleotide sequence (e.g., the first nucleotide sequence is identical to or substantially identical to a nucleotide sequence of SEQ ID NO: 1-3). Genomic DNA in the subject comprises a second nucleotide sequence having one or more polymorphic variations associated with osteoarthritis (e.g., the second nucleotide sequence is identical to or substantially identical to a nucleotide sequence of SEQ ID NO: 1). The first and second nucleotide sequences typically are substantially identical to one another, and the first nucleotide sequence comprises fewer polymorphic variations associated with osteoarthritis than the second nucleotide sequence. The first nucleotide sequence may comprise a gene sequence that encodes a full-length polypeptide or a fragment thereof. The subject is often a human. Allele therapy methods often are utilized in conjunction with a method of first determining whether a subject has genomic DNA that includes polymorphic variants associated with osteoarthritis.

[0192] In another allele therapy embodiment, provided herein is a method which comprises contacting one or more cells in the subject or from the subject with a polypeptide encoded by a nucleic acid having a first nucleotide sequence (e.g., the first nucleotide sequence is identical to or substantially identical to the nucleotide sequence of SEQ ID NO: 1-3). Genomic DNA in the subject comprises a second nucleotide sequence having one or more polymorphic variations associated with osteoarthritis (e.g., the second nucleotide sequence is identical to or substantially identical to a nucleotide sequence of SEQ ID NO: 1). The first and second nucleotide sequences typically are substantially identical to one another, and the first nucleotide sequence comprises fewer polymorphic variations associated with osteoarthritis than the second nucleotide sequence. The first nucleotide sequence may comprise a gene sequence that encodes a full-length polypeptide or a fragment thereof. The subject is often a human. The method often comprises supplementing arthritis-associated *ADAMTS2* polypeptide with a non-arthritis-associated *ADAMTS2* polypeptide or fragment thereof, where the non-arthritis-associated form of

*ADAMTS2* contains an isoleucine at position 245 of SEQ ID NO: 4 having enzymatic activity. The arthritis-associated *ADAMTS2* polypeptide sometimes contains a valine at position 245 of SEQ ID NO: 4 having an altered enzymatic activity varying from the non-arthritis-associated polypeptide.

[0193] In an embodiment, provided is a method of increasing the synthesis of procollagen II comprising providing or administering to individuals in need of increasing levels of type II collagen the pharmaceutical or physiologically acceptable composition comprising active human *ADAMTS2* protein or fragment thereof, where *ADAMTS2* polypeptide fragments having activity are selected from amino acids 252-1211, 253-1211, 254-1211, 255-1211, 256-1211, 257-1211, 258-1211, 259-1211 or 260-1211 of SEQ ID NO: 4, where it is understood that the active form of *ADAMTS2* does not contain the propeptide domain.

[0194] In another embodiment, provided herein is a method of increasing the synthesis of procollagen II comprising providing or administering to individuals in need of increasing levels of type II collagen the pharmaceutical or physiologically acceptable composition comprising an enzyme or molecule capable of cleaving *ADAMTS2* propeptide, e.g., a furin-type endopeptidase or N-ethylmaleimide described herein

[0195] For antibody-based therapies, antibodies can be generated that are both specific for target molecules and that reduce target molecule activity. Such antibodies may be administered in instances where antagonizing a target molecule function is appropriate for the treatment of osteoarthritis.

[0196] In circumstances where stimulating antibody production in an animal or a human subject by injection with a target molecule is harmful to the subject, it is possible to generate an immune response against the target molecule by use of anti-idiotypic antibodies (*see, e.g., Herlyn, Ann. Med.; 31(1): 66-78 (1999); and Bhattacharya-Chatterjee & Foon, Cancer Treat. Res.; 94: 51-68 (1998)*). Introducing an anti-idiotypic antibody to a mammal or human subject often stimulates production of anti-anti-idiotypic antibodies, which typically are specific to the target molecule. Vaccines directed to osteoarthritis also may be generated in this fashion.

[0197] In instances where the target molecule is intracellular and whole antibodies are used, internalizing antibodies may be preferred. Lipofectin or liposomes can be used to deliver the antibody or a fragment of the Fab region that binds to the target antigen into cells. Where fragments of the antibody are used, the smallest inhibitory fragment that binds to the target antigen is preferred. For example, peptides having an amino acid sequence corresponding to the Fv region of the antibody can be used. Alternatively, single chain neutralizing antibodies that bind to intracellular target antigens can also be administered. Such single chain antibodies can be administered, for example, by expressing nucleotide sequences encoding single-chain antibodies within the target cell population (*see, e.g., Marasco et al., Proc. Natl. Acad. Sci. USA 90: 7889-7893 (1993)*).

[0198] Modulators can be administered to a patient at therapeutically effective doses to treat osteoarthritis. A therapeutically effective dose refers to an amount of the modulator sufficient to result in amelioration of symptoms of osteoarthritis. Toxicity and therapeutic efficacy of modulators can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, for determining the LD<sub>50</sub> (the dose lethal to 50% of the population) and the ED<sub>50</sub> (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD<sub>50</sub>/ED<sub>50</sub>. Modulators that exhibit large therapeutic indices are preferred. While modulators that exhibit toxic side effects can be used, care should be taken to design a delivery system that targets such molecules to the site of affected tissue in order to minimize potential damage to uninfected cells, thereby reducing side effects.

[0199] Data obtained from cell culture assays and animal studies can be used in formulating a range of dosages for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED<sub>50</sub> with little or no toxicity. The dosage can vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the methods described herein, the therapeutically effective dose can be estimated initially from cell culture assays. A dose can be formulated in animal models to achieve a circulating plasma concentration range that includes the IC<sub>50</sub> (*i.e.*, the concentration of the test compound that achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma can be measured, for example, by high performance liquid chromatography.

[0200] Another example of effective dose determination for an individual is the ability to directly assay levels of “free” and “bound” compound in the serum of the test subject. Such assays may utilize antibody mimics and/or “biosensors” that have been created through molecular imprinting techniques. Molecules that modulate target molecule activity are used as a template, or “imprinting molecule”, to spatially organize polymerizable monomers prior to their polymerization with catalytic reagents. The subsequent removal of the imprinted molecule leaves a polymer matrix which contains a repeated “negative image” of the compound and is able to selectively rebind the molecule under biological assay conditions. A detailed review of this technique can be seen in Ansell *et al.*, *Current Opinion in Biotechnology* 7: 89-94 (1996) and in Shea, *Trends in Polymer Science* 2: 166-173 (1994). Such “imprinted” affinity matrixes are amenable to ligand-binding assays, whereby the immobilized monoclonal antibody component is replaced by an appropriately imprinted matrix. An example of the use of such matrixes in this way can be seen in Vlatakis, *et al.*, *Nature* 361: 645-647 (1993). Through the use of isotope-labeling, the “free” concentration of compound which modulates target molecule expression or activity readily can be monitored and used in calculations of IC<sub>50</sub>. Such “imprinted” affinity matrixes can also be designed to include fluorescent groups whose photon-emitting properties

measurably change upon local and selective binding of target compound. These changes readily can be assayed in real time using appropriate fiberoptic devices, in turn allowing the dose in a test subject to be quickly optimized based on its individual  $IC_{50}$ . An example of such a “biosensor” is discussed in Kriz *et al.*, *Analytical Chemistry* 67: 2142-2144 (1995).

[0201] The examples set forth below are intended to illustrate but not limit the invention.

#### Examples

[0202] In the following studies a group of subjects was selected according to specific parameters relating to osteoarthritis. Nucleic acid samples obtained from individuals in the study group were subjected to genetic analysis, which identified associations between osteoarthritis and a polymorphism in the *ADAMTS2* gene on chromosome five. The polymorphism was genotyped again in two replication cohorts consisting of individuals selected for OA. In addition, SNPs proximal to the incident polymorphism were identified and allelotyped in OA case and control pools. Methods are described for producing *ADAMTS2* polypeptide and *ADAMTS2* polypeptide variants *in vitro* or *in vivo*, *ADAMTS2* nucleic acids or polypeptides and variants thereof are utilized for screening test molecules for those that interact with *ADAMTS2* molecules. Test molecules identified as interactors with *ADAMTS2* molecules and *ADAMTS2* variants are further screened *in vivo* to determine whether they treat osteoarthritis.

#### Example 1

##### Samples and Pooling Strategies

##### Sample Selection

[0203] Blood samples were collected from individuals diagnosed with knee osteoarthritis, which were referred to as case samples. Also, blood samples were collected from individuals not diagnosed with knee osteoarthritis as gender and age-matched controls. A database was created that listed all phenotypic trait information gathered from individuals for each case and control sample. Genomic DNA was extracted from each of the blood samples for genetic analyses.

##### DNA Extraction from Blood Samples

[0204] Six to ten milliliters of whole blood was transferred to a 50 ml tube containing 27 ml of red cell lysis solution (RCL). The tube was inverted until the contents were mixed. Each tube was incubated for 10 minutes at room temperature and inverted once during the incubation. The tubes were then centrifuged for 20 minutes at 3000 x g and the supernatant was carefully poured off. 100-200  $\mu$ l of residual liquid was left in the tube and was pipetted repeatedly to resuspend the pellet in the residual supernatant. White cell lysis solution (WCL) was added to the tube and pipetted repeatedly until

completely mixed. While no incubation was normally required, the solution was incubated at 37°C or room temperature if cell clumps were visible after mixing until the solution was homogeneous. 2 ml of protein precipitation was added to the cell lysate. The mixtures were vortexed vigorously at high speed for 20 sec to mix the protein precipitation solution uniformly with the cell lysate, and then centrifuged for 10 minutes at 3000 x g. The supernatant containing the DNA was then poured into a clean 15 ml tube, which contained 7 ml of 100% isopropanol. The samples were mixed by inverting the tubes gently until white threads of DNA were visible. Samples were centrifuged for 3 minutes at 2000 x g and the DNA was visible as a small white pellet. The supernatant was decanted and 5 ml of 70% ethanol was added to each tube. Each tube was inverted several times to wash the DNA pellet, and then centrifuged for 1 minute at 2000 x g. The ethanol was decanted and each tube was drained on clean absorbent paper. The DNA was dried in the tube by inversion for 10 minutes, and then 1000 µl of 1X TE was added. The size of each sample was estimated, and less TE buffer was added during the following DNA hydration step if the sample was smaller. The DNA was allowed to rehydrate overnight at room temperature, and DNA samples were stored at 2-8°C.

[0205] DNA was quantified by placing samples on a hematology mixer for at least 1 hour. DNA was serially diluted (typically 1:80, 1:160, 1:320, and 1:640 dilutions) so that it would be within the measurable range of standards. 125 µl of diluted DNA was transferred to a clear U-bottom microtitre plate, and 125 µl of 1X TE buffer was transferred into each well using a multichannel pipette. The DNA and 1X TE were mixed by repeated pipetting at least 15 times, and then the plates were sealed. 50 µl of diluted DNA was added to wells A5-H12 of a black flat bottom microtitre plate. Standards were inverted six times to mix them, and then 50 µl of 1X TE buffer was pipetted into well A1, 1000 ng/ml of standard was pipetted into well A2, 500 ng/ml of standard was pipetted into well A3, and 250 ng/ml of standard was pipetted into well A4. PicoGreen (Molecular Probes, Eugene, Oregon) was thawed and freshly diluted 1:200 according to the number of plates that were being measured. PicoGreen was vortexed and then 50 µl was pipetted into all wells of the black plate with the diluted DNA. DNA and PicoGreen were mixed by pipetting repeatedly at least 10 times with the multichannel pipette. The plate was placed into a Fluoroskan Ascent Machine (microplate fluorometer produced by Labsystems) and the samples were allowed to incubate for 3 minutes before the machine was run using filter pairs 485 nm excitation and 538 nm emission wavelengths. Samples having measured DNA concentrations of greater than 450 ng/µl were re-measured for confirmation. Samples having measured DNA concentrations of 20 ng/µl or less were re-measured for confirmation.

Pooling Strategies – Discovery Cohort

[0206] Samples were derived from the Nottingham knee OA family study (UK) where index cases were identified through a knee replacement registry. Siblings were approached and assessed with knee x-rays and assigned status as affected or unaffected. In all 1,157 individuals were available. In order to create same-sex pools of appropriate sizes, 335 unrelated female individuals with OA from the Nottingham OA sample were selected for the case pool. The control pool was made up of unrelated female individuals from the St. Thomas twin study (England) with normal knee x-rays and without other indications of OA, regardless of anatomical location, as well as lacking family history of OA. The St. Thomas twin study consists of Caucasian, female participants from the St. Thomas' Hospital, London, adult-twin registry, which is a voluntary registry of >4,000 twin pairs ranging from 18 to 76 years of age. The female case samples and female control samples are described further in Table 1 below.

[0207] A select set of samples from each group were utilized to generate pools, and one pool was created for each group. Each individual sample in a pool was represented by an equal amount of genomic DNA. For example, where 25 ng of genomic DNA was utilized in each PCR reaction and there were 200 individuals in each pool, each individual would provide 125 pg of genomic DNA. Inclusion or exclusion of samples for a pool was based upon the following criteria: the sample was derived from an individual characterized as Caucasian; the sample was derived from an individual of British paternal and maternal descent; case samples were derived from individuals diagnosed with specific knee osteoarthritis (OA) and were recruited from an OA knee replacement clinic. Control samples were derived from individuals free of OA, family history of OA, and rheumatoid arthritis. Also, sufficient genomic DNA was extracted from each blood sample for all allelotyping and genotyping reactions performed during the study. Phenotype information from each individual was collected and included age of the individual, gender, family history of OA, general medical information (e.g., height, weight, thyroid disease, diabetes, psoriasis, hysterectomy), joint history (previous and current symptoms, joint-related operations, age at onset of symptoms, date of primary diagnosis, age of individual as of primary diagnosis and order of involvement), and knee-related findings (crepitus, restricted passive movement, bony swelling/deformity). Additional knee information included knee history, current symptoms, any major knee injury, meniscectomy, knee replacement surgery, age of surgery, and treatment history (including hormone replace therapy (HRT)). Samples that met these criteria were added to appropriate pools based on disease status.



[0208] The selection process yielded the pools set forth in Table 1, which were used in the studies that follow:

TABLE 1

	Female case	Female control
<b>Pool size</b> (Number)	335	335
<b>Pool Criteria</b> (ex: case/control)	control	case
<b>Mean Age</b> (ex: years)	57.21	69.95

### Example 2

#### Association of Polymorphic Variants with Osteoarthritis

[0209] A whole-genome screen was performed to identify particular SNPs associated with occurrence of osteoarthritis. As described in Example 1, two sets of samples were utilized, which included samples from female individuals having knee osteoarthritis (osteoarthritis cases), and samples from female individuals not having knee osteoarthritis (female controls). The initial screen of each pool was performed in an allelotyping study, in which certain samples in each group were pooled. By pooling DNA from each group, an allele frequency for each SNP in each group was calculated. These allele frequencies were then compared to one another. Particular SNPs were considered as being associated with osteoarthritis when allele frequency differences calculated between case and control pools were statistically significant. SNP disease association results obtained from the allelotyping study were then validated by genotyping each associated SNP across all samples from each pool. The results of the genotyping then were analyzed, allele frequencies for each group were calculated from the individual genotyping results, and a p-value was calculated to determine whether the case and control groups had statistically significant differences in allele frequencies for a particular SNP. When the genotyping results agreed with the original allelotyping results, the SNP disease association was considered validated at the genetic level.

#### SNP Panel Used for Genetic Analyses

[0210] A whole-genome SNP screen began with an initial screen of approximately 25,000 SNPs over each set of disease and control samples using a pooling approach. The pools studied in the screen are described in Example 1. The SNPs analyzed in this study were part of a set of 25,488 SNPs confirmed as being statistically polymorphic as each is characterized as having a minor allele frequency of greater than 10%. The SNPs in the set reside in genes or in close proximity to genes, and many reside

in gene exons. Specifically, SNPs in the set are located in exons, introns, and within 5,000 base-pairs upstream of a transcription start site of a gene. In addition, SNPs were selected according to the following criteria: they are located in ESTs; they are located in Locuslink or Ensembl genes; and they are located in Genomatix promoter predictions. SNPs in the set were also selected on the basis of even spacing across the genome, as depicted in Table 2.

[0211] A case-control study design using a whole genome association strategy involving approximately 28,000 single nucleotide polymorphisms (SNPs) was employed. Approximately 25,000 SNPs were evenly spaced in gene-based regions of the human genome with a median inter-marker distance of about 40,000 base pairs. Additionally, approximately 3,000 SNPs causing amino acid substitutions in genes described in the literature as candidates for various diseases were used. The case-control study samples were of female Caucasian origin (British paternal and maternal descent) 670 individuals were equally distributed in two groups: female controls and female cases. The whole genome association approach was first conducted on 2 DNA pools representing the 2 groups. Significant markers were confirmed by individual genotyping.

**TABLE 2**

<u>General Statistics</u>		<u>Spacing Statistics</u>	
Total # of SNPs	25,488	Median	37,058 bp
# of Exonic SNPs	>4,335 (17%)	Minimum*	1,000 bp
# SNPs with refSNP ID	20,776 (81%)	Maximum*	3,000,000 bp
Gene Coverage	>10,000	Mean	122,412 bp
Chromosome Coverage	All	Std Deviation	373,325 bp
		<i>*Excludes outliers</i>	

#### Allelotyping and Genotyping Results

[0212] The genetic studies summarized above and described in more detail below identified an allelic variant in the ADAMTS2 gene that is associated with osteoarthritis.

#### Assay for Verifying, Allelotyping, and Genotyping SNPs

[0213] A MassARRAY™ system (Sequenom, Inc.) was utilized to perform SNP genotyping in a high-throughput fashion. This genotyping platform was complemented by a homogeneous, single-tube assay method (hME™ or homogeneous MassEXTEND™ (Sequenom, Inc.)) in which two genotyping primers anneal to and amplify a genomic target surrounding a polymorphic site of interest. A third primer (the MassEXTEND™ primer), which is complementary to the amplified target up to but not

including the polymorphism, was then enzymatically extended one or a few bases through the polymorphic site and then terminated.

[0214] For each polymorphism, SpectroDESIGNER™ software (Sequenom, Inc.) was used to generate a set of PCR primers and a MassEXTEND™ primer which were used to genotype the polymorphism. Other primer design software could be used or one of ordinary skill in the art could manually design primers based on his or her knowledge of the relevant factors and considerations in designing such primers. Table 3 shows PCR primers and Table 4 shows extension primers used for analyzing polymorphisms. The initial PCR amplification reaction was performed in a 5 µl total volume containing 1X PCR buffer with 1.5 mM MgCl<sub>2</sub> (Qiagen), 200 µM each of dATP, dGTP, dCTP, dTTP (Gibco-BRL), 2.5 ng of genomic DNA, 0.1 units of HotStar DNA polymerase (Qiagen), and 200 nM each of forward and reverse PCR primers specific for the polymorphic region of interest.

**TABLE 3: PCR Primers**

SNP Reference	Forward PCR primer	Reverse PCR primer
rs398829	ACGTTGGATGTAGTCATCGTCCGCAGCATG	ACGTTGGATGAAGACGGTGTCTCTCCTTG

[0215] Samples were incubated at 95°C for 15 minutes, followed by 45 cycles of 95°C for 20 seconds, 56°C for 30 seconds, and 72°C for 1 minute, finishing with a 3 minute final extension at 72°C. Following amplification, shrimp alkaline phosphatase (SAP) (0.3 units in a 2 µl volume) (Amersham Pharmacia) was added to each reaction (total reaction volume was 7 µl) to remove any residual dNTPs that were not consumed in the PCR step. Samples were incubated for 20 minutes at 37°C, followed by 5 minutes at 85°C to denature the SAP.

[0216] Once the SAP reaction was complete, a primer extension reaction was initiated by adding a polymorphism-specific MassEXTEND™ primer cocktail to each sample. Each MassEXTEND™ cocktail included a specific combination of dideoxynucleotides (ddNTPs) and deoxynucleotides (dNTPs) used to distinguish polymorphic alleles from one another. Methods for verifying, allelotyping and genotyping SNPs are disclosed, for example, in U.S. Pat. No. 6,258,538, the content of which is hereby incorporated by reference. In Table 4, ddNTPs are shown and the fourth nucleotide not shown is the dNTP.

**TABLE 4: Extension Primers**

SNP Reference	Extend Probe	Termination Mix
rs398829	TGGCGTGCTCCTCTAGGA	ACG

[0217] The MassEXTEND™ reaction was performed in a total volume of 9 µl, with the addition of 1X ThermoSequenase buffer, 0.576 units of ThermoSequenase (Amersham Pharmacia), 600 nM MassEXTEND™ primer, 2 mM of ddATP and/or ddCTP and/or ddGTP and/or ddTTP, and 2 mM of dATP or dCTP or dGTP or dTTP. The deoxy nucleotide (dNTP) used in the assay normally was complementary to the nucleotide at the polymorphic site in the amplicon. Samples were incubated at 94°C for 2 minutes, followed by 55 cycles of 5 seconds at 94°C, 5 seconds at 52°C, and 5 seconds at 72°C.

[0218] Following incubation, samples were desalted by adding 16 µl of water (total reaction volume was 25 µl), 3 mg of SpectroCLEAN™ sample cleaning beads (Sequenom, Inc.) and allowed to incubate for 3 minutes with rotation. Samples were then robotically dispensed using a piezoelectric dispensing device (SpectroJET™ (Sequenom, Inc.)) onto either 96-spot or 384-spot silicon chips containing a matrix that crystallized each sample (SpectroCHIP™ (Sequenom, Inc.)). Subsequently, MALDI-TOF mass spectrometry (Biflex and Autoflex MALDI-TOF mass spectrometers (Bruker Daltonics) can be used) and SpectroTYPER RT™ software (Sequenom, Inc.) were used to analyze and interpret the SNP genotype for each sample.

#### Genetic Analysis

[0219] Minor allelic frequencies for the polymorphisms set forth in Table A were verified as being 10% or greater using the extension assay described above in a group of samples isolated from 92 individuals originating from the state of Utah in the United States, Venezuela and France (Coriell cell repositories).

[0220] Genotyping results are shown for female pools in Table 5. In Table 5, “AF” refers to allelic frequency; and “F case” and “F control” refer to female case and female control groups, respectively.

**TABLE 5: Genotyping Results**

SNP Reference	AF F case	AF F control	p-value
rs398829	G = 0.740 A = 0.260	G = 0.652 A = 0.348	0.0002

[0221] All of the single marker alleles set forth in Table A were considered validated, since the genotyping data agreed with the allelotyping data and each SNP significantly associated with osteoarthritis. Particularly significant associations with osteoarthritis are indicated by a calculated p-value of less than 0.05 for genotype results.

### Example 3

#### Association of Polymorphic Variants with Osteoarthritis in Replication Cohorts

[0222] The single marker polymorphism set forth in Table A was genotyped again in two replication cohorts consisting of individuals selected for OA.

#### Sample Selection and Pooling Strategies – Replication Sample 1

[0223] A second case control sample (replication sample #1) was created by using 100 Caucasian female cases from Chingford, UK, and 148 unrelated female cases from the St. Thomas twin study. Cases were defined as having Kellgren-Lawrence (KL) scores of at least 2 in at least one knee x-ray. In addition, 199 male knee replacement cases from Nottingham were included. (For a cohort description, see the Nottingham description provided in Example 1). The control pool was made up of unrelated female individuals from the St. Thomas twin study (England) with normal knee x-rays and without other indications of OA, regardless of anatomical location, as well as lacking family history of OA. The St. Thomas twin study consists of Caucasian, female participants from the St. Thomas' Hospital, London, adult-twin registry, which is a voluntary registry of >4,000 twin pairs ranging from 18 to 76 years of age. The replication sample 1 cohort was used to replicate the initial results. Table 6 below summarizes the selected phenotype data collected from the case and control individuals.

**TABLE 6**

<b>Phenotype</b>	<b>Female cases (n=248): median (range)/ (n,%)</b>	<b>Male cases (n=199): median (range)/ (n,%)</b>	<b>Female controls (n=313): mean (range)/ (n,%)</b>
Age	59 (39- 73)	66 (45- 73)	55 (50- 72)
Height (cm)	162 (141- 178)	175 (152- 198)	162 (141- 176)
Weight (kg)	68 (51- 123)	86 (62- 127)	64 (40- 111)
Body mass index (kg/m <sup>2</sup> )	26 (18- 44)	29 (21- 41)	24 (18- 46)
Kellgren- Lawrence* left knee	0 (63, 26%), 1 (20, 8%), 2 (105, 43%), 3 (58, 23%), 4 (1, 0%)	NA	NA
Kellgren- Lawrence* right knee	0 (43, 7%), 1 (18, 7%), 2 (127, 52%), 3 (57, 23%), 4 (1, 0%)	NA	NA
KL* >2 both knees	No (145, 59%), Yes (101, 41%)	NA	NA
KL* >2 either knee	No (0, 0%), Yes (248, 100%)	NA	NA

\* 0: normal, 1: doubtful, 2: definite osteophyte (bony protuberance), 3: joint space narrowing (with or without osteophyte), 4: joint deformity

Sample Selection and Pooling Strategies – Replication Sample 2

[0224] A third case control sample (replication sample #2) was created by using individuals with symptoms of OA from Newfoundland, Canada. These individuals were recruited and examined by rheumatologists. Affected joints were x-rayed and a final diagnosis of definite or probable OA was made according to American College of Rheumatology criteria by a single rheumatologist to avoid any inter-examiner diagnosis variability. Controls were recruited from volunteers without any symptoms from the musculoskeletal system based on a normal joint exam performed by a rheumatologist. Only cases with a diagnosis of definite OA were included in the study. Only individuals of Caucasian origin were included. The cases consisted of 228 individuals with definite knee OA, 106 individuals with definite hip OA, and 74 individuals with hip OA.

**TABLE 7**

Phenotype	Case	Control
Age at Visit	62.7	52.5
Sex (Female/Male)	227/119	174/101
Knee OA Xray: No	35% (120)	80% (16)
Unknown	1% (4)	0% (0)
Yes	64% (221)	20% (4)
Hip OA Xray: No	63% (215)	80% (16)
Unknown	2% (7)	0% (0)
Yes	35% (121)	20% (4)

Assay for Verifying, Allelotyping, and Genotyping SNPs

[0225] Genotyping of the replication cohorts described in Tables 6 and 7 was performed using the same methods used for the original genotyping, as described herein. A MassARRAY™ system (Sequenom, Inc.) was utilized to perform SNP genotyping in a high-throughput fashion. This genotyping platform was complemented by a homogeneous, single-tube assay method (hME™ or homogeneous MassEXTEND™ (Sequenom, Inc.)) in which two genotyping primers anneal to and amplify a genomic target surrounding a polymorphic site of interest. A third primer (the MassEXTEND™ primer), which is complementary to the amplified target up to but not including the polymorphism, was then enzymatically extended one or a few bases through the polymorphic site and then terminated.

[0226] For each polymorphism, SpectroDESIGNER™ software (Sequenom, Inc.) was used to generate a set of PCR primers and a MassEXTEND™ primer which were used to genotype the polymorphism. Other primer design software could be used or one of ordinary skill in the art could

manually design primers based on his or her knowledge of the relevant factors and considerations in designing such primers. Table 3 shows PCR primers and Table 4 shows extension probes used for analyzing (*e.g.*, genotyping) polymorphisms in the replication cohorts. The initial PCR amplification reaction was performed in a 5  $\mu$ l total volume containing 1X PCR buffer with 1.5 mM MgCl<sub>2</sub> (Qiagen), 200  $\mu$ M each of dATP, dGTP, dCTP, dTTP (Gibco-BRL), 2.5 ng of genomic DNA, 0.1 units of HotStar DNA polymerase (Qiagen), and 200 nM each of forward and reverse PCR primers specific for the polymorphic region of interest.

[0227] Samples were incubated at 95°C for 15 minutes, followed by 45 cycles of 95°C for 20 seconds, 56°C for 30 seconds, and 72°C for 1 minute, finishing with a 3 minute final extension at 72°C. Following amplification, shrimp alkaline phosphatase (SAP) (0.3 units in a 2  $\mu$ l volume) (Amersham Pharmacia) was added to each reaction (total reaction volume was 7  $\mu$ l) to remove any residual dNTPs that were not consumed in the PCR step. Samples were incubated for 20 minutes at 37°C, followed by 5 minutes at 85°C to denature the SAP.

[0228] Once the SAP reaction was complete, a primer extension reaction was initiated by adding a polymorphism-specific MassEXTEND™ primer cocktail to each sample. Each MassEXTEND™ cocktail included a specific combination of dideoxynucleotides (ddNTPs) and deoxynucleotides (dNTPs) used to distinguish polymorphic alleles from one another. Methods for verifying, allelotyping and genotyping SNPs are disclosed, for example, in U.S. Pat. No. 6,258,538, the content of which is hereby incorporated by reference. In Table 7, ddNTPs are shown and the fourth nucleotide not shown is the dNTP.

[0229] The MassEXTEND™ reaction was performed in a total volume of 9  $\mu$ l, with the addition of 1X ThermoSequenase buffer, 0.576 units of ThermoSequenase (Amersham Pharmacia), 600 nM MassEXTEND™ primer, 2 mM of ddATP and/or ddCTP and/or ddGTP and/or ddTTP, and 2 mM of dATP or dCTP or dGTP or dTTP. The deoxy nucleotide (dNTP) used in the assay normally was complementary to the nucleotide at the polymorphic site in the amplicon. Samples were incubated at 94°C for 2 minutes, followed by 55 cycles of 5 seconds at 94°C, 5 seconds at 52°C, and 5 seconds at 72°C.

[0230] Following incubation, samples were desalted by adding 16  $\mu$ l of water (total reaction volume was 25  $\mu$ l), 3 mg of SpectroCLEAN™ sample cleaning beads (Sequenom, Inc.) and allowed to incubate for 3 minutes with rotation. Samples were then robotically dispensed using a piezoelectric dispensing device (SpectroJET™ (Sequenom, Inc.)) onto either 96-spot or 384-spot silicon chips containing a matrix that crystallized each sample (SpectroCHIP™ (Sequenom, Inc.)). Subsequently, MALDI-TOF mass spectrometry (Biflex and Autoflex MALDI-TOF mass spectrometers (Bruker Daltonics) can be used) and

SpectroTYPER RT™ software (Sequenom, Inc.) were used to analyze and interpret the SNP genotype for each sample.

#### Genetic Analysis

[0231] Genotyping results for replication cohorts #1 and #2 are provided in Tables 8 and 9, respectively.

**TABLE 8**

rsID	Replication #1 (Mixed Male/Female cases and Female controls)				Meta-analysis Disc. + Rep #1
	AF OA Con	AF OA Cas	Delta	P-value	P-value
rs398829	0.30	0.28	0.02	0.307	0.0260

**TABLE 9**

rsID	Replication #2 (Newfoundland) (Male/Female cases and controls)				Meta-analysis Disc. + Rep #2
	AF OA Con	AF OA Cas	Delta	P-value	Not Done
rs398829	0.27	0.28	-0.013	0.627	

[0232] To combine the evidence for association from multiple sample collections, a meta-analysis procedure was employed. The allele frequencies were compared between cases and controls within the discovery sample, as well as within the replication cohort #1 using the DerSimonian-Laird approach (DerSimonian, R. and N. Laird. 1986. Meta-analysis in clinical trials. Control Clin Trials 7: 177-188.)

[0233] The absence of a statistically significant association in one or more of the replication cohorts should not be interpreted as minimizing the value of the original finding. There are many reasons why a biologically derived association identified in a sample from one population would not replicate in a sample from another population. The most important reason is differences in population history. Due to bottlenecks and founder effects, there may be common disease predisposing alleles present in one population that are relatively rare in another, leading to a lack of association in the candidate region. Also, because common diseases such as arthritis-related disorders are the result of susceptibilities in many genes and many environmental risk factors, differences in population-specific genetic and environmental backgrounds could mask the effects of a biologically relevant allele. For these and other reasons, statistically strong results in the original, discovery sample that did not replicate in one or more of the replication samples may be further evaluated in additional replication cohorts and experimental systems.



#### Example 4

##### ADAMTS2 Region Proximal SNPs

[0234] It has been discovered that SNP rs398829 in *ADAMTS2* is associated with occurrence of osteoarthritis in subjects. This gene encodes a disintegrin and metalloproteinase with thrombospondin motifs-2 (*ADAMTS2*), which is a member of the *ADAMTS* protein family. Members of the family share several distinct protein modules, including a propeptide region, a metalloproteinase domain, a disintegrin-like domain, and a thrombospondin type 1 (TS) motif. *ADAMTS2* is involved in collagens 1, 2 and 5 N-terminal processing, (type II collagen is the major form in cartilage). Mutations in this gene cause Ehlers-Danlos syndrome type VIIC, a recessively inherited connective-tissue disorder that causes loose joints and fragile skin. Mild loss of function may exacerbate physical joint damage leading to a predisposition to OA and incorrectly processed collagen can act dominantly to inhibit self assembly of fibrils. Alternative splicing of the gene generates 2 transcript variants. The short transcript encodes a protein, which has no significant procollagen N-peptidase activity.

[0235] Two hundred-nine additional allelic variants proximal to rs398829 were identified and subsequently allelotyped in osteoarthritis case and control sample sets as described in Examples 1 and 2. The polymorphic variants are set forth in Table 10. The chromosome positions provided in column four of Table 10 are based on Genome “Build 34” of NCBI’s GenBank.

**TABLE 10**

dbSNP rs#	Chromosome	Position in SEQ ID NO: 1	Chromosome Position	Allele Variants
rs2278221	5	210	178695460	c/t
rs1650358	5	3608	178698858	c/g
rs1643818	5	3609	178698859	c/g
rs3733916	5	4318	178699568	c/t
rs1624933	5	5593	178700843	a/g
rs1624857	5	5629	178700879	c/t
rs1624832	5	5639	178700889	a/g
rs1624829	5	5640	178700890	c/t
rs2161171	5	8943	178704193	a/c
rs1530499	5	17968	178713218	a/g
rs888764	5	19887	178715137	a/g
rs873987	5	21034	178716284	a/g
rs4078699	5	21085	178716335	c/t
rs870311	5	21596	178716846	a/g
rs1643817	5	23379	178718629	a/c
rs1643816	5	23432	178718682	a/c
rs1650355	5	24007	178719257	a/c
rs888763	5	26121	178721371	a/g
rs1862212	5	26273	178721523	a/t

dbSNP rs#	Chromosome	Position in SEQ ID NO: 1	Chromosome Position	Allele Variants
rs1110514	5	26755	178722005	a/t
rs3797600	5	27411	178722661	c/t
rs3797602	5	27710	178722960	g/t
rs3797603	5	27842	178723092	c/t
rs3776819	5	28379	178723629	c/t
rs252076	5	29603	178724853	c/t
rs252075	5	31232	178726482	c/g
rs252074	5	31504	178726754	a/g
rs252068	5	32583	178727833	c/g
rs252069	5	32794	178728044	a/g
rs194040	5	32840	178728090	c/t
rs252070	5	33044	178728294	c/t
rs3797606	5	33150	178728400	a/c
rs171667	5	33218	178728468	a/g
rs187539	5	33513	178728763	c/t
rs3836834	5	33959	178729209	- /tatcaaactaccatga aa
rs252071	5	34486	178729736	a/g
rs252072	5	36289	178731539	c/t
rs252073	5	36570	178731820	c/t
rs379589	5	38247	178733497	a/t
rs2052472	5	38477	178733727	a/c
rs2052471	5	38518	178733768	c/t
rs2052470	5	38529	178733779	c/t
rs2052469	5	38667	178733917	a/g
rs3797608	5	39781	178735031	c/t
rs3797609	5	39856	178735106	c/t
rs3822601	5	39927	178735177	c/t
rs153131	5	40506	178735756	a/g
rs751546	5	41869	178737119	c/g
rs2279979	5	42452	178737702	c/t
rs252060	5	44788	178740038	c/t
rs3797610	5	46059	178741309	a/c
rs194039	5	46846	178742096	a/g
rs168773	5	47712	178742962	a/t
rs252061	5	48796	178744046	c/t
rs187537	5	49441	178744691	c/g
rs252062	5	49602	178744852	a/t
rs2431255	5	49723	178744973	a/c
rs3797612	5	50050	178745300	c/t
rs3797613	5	50171	178745421	c/t
rs614114	5	50477	178745727	c/t
rs252063	5	50818	178746068	c/t
rs252064	5	50833	178746083	c/t
rs252065	5	50881	178746131	a/g
rs450502	5	50882	178746132	a/g
rs439252	5	51386	178746636	c/t

dbSNP rs#	Chromosome	Position in SEQ ID NO: 1	Chromosome Position	Allele Variants
rs252066	5	51534	178746784	c/t
rs457957	5	52317	178747567	a/g
rs3797614	5	52368	178747618	c/t
rs423552	5	52970	178748220	a/g
rs398829	5	53023	178748273	a/g
rs416646	5	53356	178748606	a/g
rs187450	5	53882	178749132	g/t
rs337807	5	54553	178749803	c/t
rs337806	5	55475	178750725	a/c
rs1396438	5	55530	178750780	a/g
rs1396437	5	55691	178750941	c/t
rs2411811	5	55848	178751098	a/c
rs2898813	5	55879	178751129	c/g
rs189256	5	56316	178751566	a/g
rs173072	5	56911	178752161	a/c
rs337805	5	57320	178752570	a/g
rs191415	5	57391	178752641	c/t
rs180045	5	57437	178752687	c/t
rs189255	5	57478	178752728	c/g
rs652766	5	57500	178752750	c/t
rs466750	5	59111	178754361	g/t
rs442406	5	59333	178754583	a/g
rs662407	5	59715	178754965	a/g
rs592971	5	59804	178755054	a/g
rs457187	5	59851	178755101	a/g
rs459490	5	59929	178755179	c/t
rs459668	5	60052	178755302	c/t
rs462646	5	60240	178755490	c/t
rs458272	5	60359	178755609	g/t
rs463455	5	60381	178755631	a/g
rs675880	5	60456	178755706	c/t
rs810617	5	60724	178755974	c/g
rs464156	5	60875	178756125	c/t
rs458083	5	60968	178756218	a/g
rs467333	5	60978	178756228	c/g
rs465381	5	60998	178756248	c/t
rs466363	5	61557	178756807	c/t
rs2457099	5	62091	178757341	c/t
rs463901	5	62645	178757895	c/t
rs465621	5	62943	178758193	a/c
rs463724	5	63131	178758381	a/t
rs465242	5	63145	178758395	g/t
rs467419	5	63406	178758656	a/g
rs456135	5	63427	178758677	c/g
rs464536	5	63554	178758804	c/t
rs461898	5	63661	178758911	a/g
rs389558	5	64093	178759343	a/g

dbSNP rs#	Chromosome	Position in SEQ ID NO: 1	Chromosome Position	Allele Variants
rs466752	5	64153	178759403	c/t
rs455655	5	64409	178759659	c/g
rs463435	5	64544	178759794	c/t
rs2174971	5	65257	178760507	c/t
rs1979979	5	65626	178760876	a/g
rs411804	5	65739	178760989	a/g
rs1623885	5	66392	178761642	c/t
rs1643811	5	66720	178761970	c/t
rs434430	5	69177	178764427	a/t
rs187538	5	69336	178764586	g/t
rs252067	5	69636	178764886	a/g
rs459319	5	69823	178765073	a/g
rs467289	5	69928	178765178	c/t
rs462644	5	70547	178765797	c/t
rs458752	5	70633	178765883	c/t
rs708320	5	71805	178767055	a/c
rs457954	5	72181	178767431	c/g
rs2411810	5	72200	178767450	c/t
rs3084687	5	72474	178767724	-/at
rs69638	5	72567	178767817	c/g
rs455452	5	72973	178768223	a/g
rs464850	5	73468	178768718	a/g
rs431472	5	73889	178769139	a/g
rs2411809	5	75730	178770980	c/t
rs2457094	5	75970	178771220	a/g
rs2457095	5	76114	178771364	a/g
rs2261740	5	76342	178771592	c/t
rs1109180	5	76449	178771699	a/g
rs1109179	5	76465	178771715	c/t
rs1109178	5	76791	178772041	a/c
rs456909	5	78042	178773292	a/g
rs469124	5	80758	178776008	a/g
rs468039	5	80778	178776028	c/t
rs467017	5	81356	178776606	a/c
rs469290	5	81576	178776826	a/g
rs469090	5	81689	178776939	c/t
rs469568	5	81759	178777009	g/t
rs468386	5	81950	178777200	c/g
rs469349	5	82562	178777812	a/c
rs469099	5	83591	178778841	c/t
rs456868	5	83700	178778950	a/g
rs465389	5	83821	178779071	c/g
rs463892	5	83842	178779092	c/g
rs468548	5	83923	178779173	g/t
rs654612	5	83929	178779179	a/c
rs468542	5	84021	178779271	c/g
rs469262	5	84175	178779425	c/t

dbSNP rs#	Chromosome	Position in SEQ ID NO: 1	Chromosome Position	Allele Variants
rs708323	5	84417	178779667	a/g
rs469089	5	84747	178779997	c/g
rs469396	5	85746	178780996	c/g
rs468723	5	86129	178781379	c/t
rs467604	5	86335	178781585	a/g
rs338874	5	87315	178782565	c/g
rs338875	5	87648	178782898	a/g
rs1385803	5	87764	178783014	a/c
rs1385804	5	87770	178783020	c/g
rs338876	5	88221	178783471	c/t
rs189803	5	90474	178785724	a/c
rs452215	5	91148	178786398	g/t
rs641170	5	91150	178786400	g/t
rs584398	5	91160	178786410	g/t
rs385330	5	91733	178786983	c/t
rs429538	5	91772	178787022	a/c
rs371229	5	91785	178787035	c/t
rs460874	5	93140	178788390	a/t
rs646121	5	93148	178788398	a/t
rs468262	5	96080	178791330	a/g
rs467863	5	96157	178791407	c/g
rs191434	5	96313	178791563	a/c
rs2054782	5	96759	178792009	c/t
rs468499	5	97026	178792276	a/c
rs180287	5	97320	178792570	c/g
rs338877	5	97732	178792982	a/t
rs650665	5	98713	178793963	c/g
rs193419	5	99707	178794957	a/c
rs180288	5	99959	178795209	c/g
rs186834	5	100009	178795259	a/g
rs189266	5	100020	178795270	c/g
rs189267	5	100065	178795315	a/c
rs170937	5	100086	178795336	c/g
rs463263	5	101270	178796520	c/g
rs463262	5	101276	178796526	g/t
rs460454	5	101371	178796621	c/t
rs460455	5	101376	178796626	c/g
rs460505	5	101439	178796689	c/t
rs931316	5	101820	178797070	c/t
rs463431	5	102392	178797642	c/g
rs461542	5	102602	178797852	a/g
rs463557	5	102604	178797854	a/c
rs191453	5	102896	178798146	c/t
rs2271212	5	189104	178884354	c/t
rs462009	5	189134	178884384	c/t
rs2271211	5	189205	178884455	a/g
rs396474	5	Not mapped	Not mapped	a/c

dbSNP rs#	Chromosome	Position in SEQ ID NO: 1	Chromosome Position	Allele Variants
rs428901	5	Not mapped	Not mapped	a/t
rs452300	5	Not mapped	Not mapped	g/t
rs670256	5	Not mapped	Not mapped	g/t

#### Assay for Verifying and Allelotyping SNPs

[0236] The methods used to verify and allelotype the 209 proximal SNPs of Table 10 are the same methods described in Examples 1 and 2 herein. The primers and probes used in these assays are provided in Table 11 and Table 12, respectively.

**TABLE 11**

dbSNP rs#	Forward PCR primer	Reverse PCR primer
rs2278221	ACGTTGGATGTCTCATGGGCCACCACAAAC	ACGTTGGATGTATGCTCCTGTCACCGGCAT
rs1650358	ACGTTGGATGTGGATGGCTCCATGTTCTTG	ACGTTGGATGAAGTGCTGGGATTACAGGTG
rs1643818	ACGTTGGATGCTGGGATTACAGGTGTGAAC	ACGTTGGATGTGGATGGCTCCATGTTCTTG
rs3733916	ACGTTGGATGCCGAGCAGGCTGTAGTGTTG	ACGTTGGATGCTTTGTACCACCTGGAACAG
rs1624933	ACGTTGGATGAGGCTGGTCTCAAACCTCCTG	ACGTTGGATGTAACAAAAGTTGGCCGTGC
rs1624857	ACGTTGGATGTGAGGTCAGGAGTTTGAGAC	ACGTTGGATGGCCACCAAGCCAGACTAAGT
rs1624832	ACGTTGGATGTGAGGTCAGGAGTTTGAGAC	ACGTTGGATGGCCACCAAGCCAGACTAAGT
rs1624829	ACGTTGGATGTGAGGTCAGGAGTTTGAGAC	ACGTTGGATGGCCACCAAGCCAGACTAAGT
rs2161171	ACGTTGGATGCCCGTCACCACCTTTATTTCC	ACGTTGGATGAGAGTGAGTCCAGTCTGCAG
rs1530499	ACGTTGGATGACTCCAAGATTTCCCATTTT	ACGTTGGATGTTCTGTGTTCCACCCTATGG
rs888764	ACGTTGGATGTAGTTGAATGTTGTATTGGC	ACGTTGGATGACCGTGATAAACACAGAATG
rs873987	ACGTTGGATGGCTGTTAATCATGTGTCGGG	ACGTTGGATGATTGGCCACATCACCAGAC
rs4078699	ACGTTGGATGGTACCGTGGATTCTTTTAGG	ACGTTGGATGGTATTGGAAGAGCAGAGAC
rs870311	ACGTTGGATGTCAGGGCTCCAGTGTTGAAG	ACGTTGGATGAAAAGGAGGAGTGCCCTGTG
rs1643817	ACGTTGGATGATGGGAACTCCTGGTCCTG	ACGTTGGATGAAAATGCAAGCCGCCACCTG
rs1643816	ACGTTGGATGTTTTCTCCCCTTTCTAGCCC	ACGTTGGATGTTGGCATGAGAGATGGACAG
rs1650355	ACGTTGGATGTCAACAGCAACAAAACCAA	ACGTTGGATGTAAATAGGTCAGAGGGTTG
rs888763	ACGTTGGATGAAGAGGAAGAGACATACCAG	ACGTTGGATGAACAACATGGACTCAGGCTG
rs1862212	ACGTTGGATGGGCCACATTTTAAACAAGGG	ACGTTGGATGTCCCCTGAGGTTTCTATAAG
rs1110514	ACGTTGGATGTGCCACGTTCCATGTTTCAG	ACGTTGGATGATCACTGTAGCCCTTCTCTG
rs3797600	ACGTTGGATGCCCTTCTGTACCTCCTTTG	ACGTTGGATGGGAAGTGACTGCTGAGCTG
rs3797602	ACGTTGGATGAGAAACAGGGACTGGCTGTGT	ACGTTGGATGAGCAGGCTCCGGGAAGTATG
rs3797603	ACGTTGGATGCACCCATCCATCATGATGTC	ACGTTGGATGTGCTACCTCAAAACAGTGGG
rs3776819	ACGTTGGATGCAAGCACCATTCCATTGCAC	ACGTTGGATGAAATGAGGATTGCAGTCCCC
rs252076	ACGTTGGATGACTTCTGACTTCAGGTGATC	ACGTTGGATGTATAGGAACGAAAGAAAGCC
rs252075	ACGTTGGATGTGGGAGCATTTGCAGGCATG	ACGTTGGATGAAGCCTCAGATGGTTCGGAG
rs252074	ACGTTGGATGTTGCGATGGCCTCCTGGCT	ACGTTGGATGAAGTTGAGGGCTCCGGAGCA
rs252068	ACGTTGGATGGGGTAGGAAGGGTTTAAGC	ACGTTGGATGGCAGCCCCTCAATTCTTTAG
rs252069	ACGTTGGATGTGCCCATTTTCTGTTATTCC	ACGTTGGATGTTTGACTTGCCGTGCAACT
rs194040	ACGTTGGATGTGCCCATTTTCTGTTATTCC	ACGTTGGATGTTTGACTTGCCGTGCAACT
rs252070	ACGTTGGATGCTCAAGGACATTGTCCCTGG	ACGTTGGATGGGAGAAGCAGCTCTCCTTTC

dbSNP rs#	Forward PCR primer	Reverse PCR primer
rs3797606	ACGTTGGATGGTTTCCCCAAACAAGAGAGC	ACGTTGGATGGGAAATGTTCAAAGCCGCAG
rs171667	ACGTTGGATGGGGAAACACATTGTAATGCG	ACGTTGGATGCCTTCCTCATTGTCTATTCC
rs187539	ACGTTGGATGAGCCACCCCAACCTTCAGGA	ACGTTGGATGTTGCTCCTGGACATGGTTTT
rs3836834	ACGTTGGATGAAGAAACGTGACTCTTGCTC	ACGTTGGATGTAGTAATTCTGATCCTGGCC
rs252071	ACGTTGGATGGCTTCAACCTGAAACAACCC	ACGTTGGATGGGGATATTCCCCTCTGAG
rs252072	ACGTTGGATGTTGTTTCCCCAAAGGCGACG	ACGTTGGATGTGTGTTTTCCAGAGCTGGAG
rs252073	ACGTTGGATGGGGAAAGGCCGAGAAAAGTC	ACGTTGGATGACAAGCTCAGCAGAGTTCCA
rs379589	ACGTTGGATGAAACACGGGAGTACTGAGCA	ACGTTGGATGTTGTTAGCTGTCTGTCCGTC
rs2052472	ACGTTGGATGAACCAGCTCAAGGATCACCC	ACGTTGGATGAAAGGAGACGGTCAGCTGTC
rs2052471	ACGTTGGATGACAGCTGACCGTCTCCTTTG	ACGTTGGATGCCCCGTCTGGACAAGCTTTT
rs2052470	ACGTTGGATGACAGCTGACCGTCTCCTTTG	ACGTTGGATGCCCCGTCTGGACAAGCTTTT
rs2052469	ACGTTGGATGAGGGAAAGATATCGCACGCG	ACGTTGGATGAGTGAACAACCTGCTCGCCTC
rs3797608	ACGTTGGATGTGCTTTGCCTTGCTTCTGC	ACGTTGGATGTGCACTAAGGGAGTGAGTGG
rs3797609	ACGTTGGATGTGCAGAAGCCAAGGCAAAGC	ACGTTGGATGACAGCATTTGGAGTCCCCTG
rs3822601	ACGTTGGATGAGGTCAGTGAGGCCTGAGAT	ACGTTGGATGTGTCTGGCCTGAAGATCGAG
rs153131	ACGTTGGATGTAATCACGTGTCCTGATCCC	ACGTTGGATGAGCTGTCTCAGTCATGTTT
rs751546	ACGTTGGATGTCCTGCTCTGCCGTTCTACA	ACGTTGGATGATCAGCTCAAAGGACCGGTG
rs2279979	ACGTTGGATGTATTGCTACCAGGAACACGTA	ACGTTGGATGAAAAAGGGGCCACTTCAGGG
rs252060	ACGTTGGATGTGGCCAGAGCCCGTGTTC	ACGTTGGATGCGGCCAATCCCCTCTCTATG
rs3797610	ACGTTGGATGAAAAGCTTCTCCCTTGGGTG	ACGTTGGATGCAAGTAGGGCAGAACTCAG
rs194039	ACGTTGGATGAAAGTGCTGGGATTACAGGC	ACGTTGGATGTGCTGGGAGAAGACATTAC
rs168773	ACGTTGGATGTGGTGCCTGAGATATCAC	ACGTTGGATGGATCCCTATCCTACCTCTTC
rs252061	ACGTTGGATGTGTACACTCCTCTGTGAAG	ACGTTGGATGCTGTCTCTCCATGCTTTTGC
rs187537	ACGTTGGATGCGAGGATGTCATGCTAAGTG	ACGTTGGATGGTACCTCGCATAAGTGGATC
rs252062	ACGTTGGATGAAGCACATTATGTGGCTGG	ACGTTGGATGCTGAAACTCAATGGGCACAG
rs2431255	ACGTTGGATGGGTGAAGACGGTGACTTATG	ACGTTGGATGCTGGTGTCTTGAAGAAGT
rs3797612	ACGTTGGATGAGTGAGGACGCAGGGCATT	ACGTTGGATGAGCGTGGGCGAGGGAGATAA
rs3797613	ACGTTGGATGATCAGAGGCAGAGACCCCC	ACGTTGGATGGGGTGTCTCTGCAGAGCGG
rs614114	ACGTTGGATGGGTTGGAGGATGTCTAGAAC	ACGTTGGATGGGCTGGATCACTAGGGTTTG
rs252063	ACGTTGGATGTTGGAATTACAGTCCGATGG	ACGTTGGATGCTGAGAGACTGAAAAGCACA
rs252064	ACGTTGGATGCTGAGAGACTGAAAAGCACA	ACGTTGGATGTTGGAATTACAGTCCGATGG
rs252065	ACGTTGGATGAAAACCTAAGGCTCAGAGGAC	ACGTTGGATGTGGGCTTGGGAATTACAGTC
rs450502	ACGTTGGATGATGAGAAAACCAAGGCTCAG	ACGTTGGATGCTGGGCTTGGGAATTACAGTC
rs439252	ACGTTGGATGATCTCTGACCTCGTGATCC	ACGTTGGATGTCATAATAACGGCCGGGTGC
rs252066	ACGTTGGATGTTTCTCTTGACCGGTCTTG	ACGTTGGATGTAAACGAATTCTGCCGATG
rs457957	ACGTTGGATGTTACGTGCATTAGAGCGAG	ACGTTGGATGAATTCCTCCCAATTCTCTC
rs3797614	ACGTTGGATGACTGCGAGCTTTAAGGAGGG	ACGTTGGATGCCAAACAGAAAGCCCTTTTC
rs423552	ACGTTGGATGGCAGGACCTCGATGTTGTAG	ACGTTGGATGATCCTAGAGGAGCACGCCAAC
rs398829	ACGTTGGATGTAGTCATCGTCCGCAGCATG	ACGTTGGATGAAGACGGTGTCTCTCTCTTG
rs416646	ACGTTGGATGGCTGGGTCTCTCACAGTCTC	ACGTTGGATGAGACAGGCACCTCTGTGACTT
rs187450	ACGTTGGATGAGAAGGCAGGGACGATATCC	ACGTTGGATGACCAAGATGAACCCCTCTGT
rs337807	ACGTTGGATGTCACCCAGTGCTGACAGCAG	ACGTTGGATGATGCTGGGATGCCATGGGTC
rs337806	ACGTTGGATGAATTAAGAGATGGGGCCACC	ACGTTGGATGGCCCTGTGTGTTTTGTCTCC
rs1396438	ACGTTGGATGTACCTTCTGGTGCCAGAATG	ACGTTGGATGCCTGGAGACAAAACACACAG
rs1396437	ACGTTGGATGTAAAACTCTGCCTGCTCGG	ACGTTGGATGTCCAGACATTCCCCGTAGGA
rs2411811	ACGTTGGATGGAGGGATGCTCTAGAACATA	ACGTTGGATGCTGAATTTACCTGAAATGG
rs2898813	ACGTTGGATGTCCTCACCCACTTTGCCTTT	ACGTTGGATGATCGTGATAATTTGGGGTG

dbSNP rs#	Forward PCR primer	Reverse PCR primer
rs189256	ACGTTGGATGCTCCCTATAGCAAGGCTCTA	ACGTTGGATGTTAACCCAGGCCATGAAGAG
rs173072	ACGTTGGATGAGCTGGAGATCTCTTGCTC	ACGTTGGATGCTAAAACAGGATGGCTCTGG
rs337805	ACGTTGGATGGAACAAACCAAGGAGCAGG	ACGTTGGATGATGTGGACAACGTTGGACTC
rs191415	ACGTTGGATGAATTACATGACTCGGACAAG	ACGTTGGATGTGCTGGTGAAGTACAGAAGG
rs180045	ACGTTGGATGGTCCCAGGTTTTCTGTTCTC	ACGTTGGATGTGTACTTCACCAGCACTGAG
rs189255	ACGTTGGATGAGGTTGCAGACTCAGTCCCA	ACGTTGGATGGGGTGATTTGCGGGAATGAG
rs652766	ACGTTGGATGGGGTGATTTGCGGGAATGAG	ACGTTGGATGACCATCCCACGATGCTCCC
rs466750	ACGTTGGATGTATCTCCTTAAATGCCTTGG	ACGTTGGATGTGACCAGGAGGAGTTAAAC
rs442406	ACGTTGGATGTGACAAGGTCACGTGTTCTG	ACGTTGGATGCCAGACAAGTCTGATACAGC
rs662407	ACGTTGGATGCCACAGTCACCACTACTGAG	ACGTTGGATGCTTGAGCCATGAGTGGAAATG
rs592971	ACGTTGGATGGGAAGCATTTCTTTGACTGC	ACGTTGGATGATTCCATCTCATGGCTCAAG
rs457187	ACGTTGGATGTGTGAGATGAGGAGTATCTG	ACGTTGGATGGCAGTCAAAGAAATGCTTCC
rs459490	ACGTTGGATGACAGATACTCCTCATCTCAC	ACGTTGGATGGGGAGTTTTGCTGTTATAGC
rs459668	ACGTTGGATGGCTTCATTTACTGAGGTCTTC	ACGTTGGATGTGAATGTTCAACGACTACAC
rs462646	ACGTTGGATGCAATTATTCGACGGAGATTA	ACGTTGGATGCTCCTCCAAATGAATCAAGAA
rs458272	ACGTTGGATGATGCCTCCTCATTGTCATT	ACGTTGGATGCCCAACAAAGTGATTCCAAC
rs463455	ACGTTGGATGATGCCTCCTCATTGTCATT	ACGTTGGATGCCCAACAAAGTGATTCCAAC
rs675880	ACGTTGGATGCAGCTCCATTGATCTGTTTC	ACGTTGGATGAAGAATGACAATGAGGAGGC
rs810617	ACGTTGGATGTGATCTCAGCTTACCACAGC	ACGTTGGATGATGCCTGTAATCCCAGCTAC
rs464156	ACGTTGGATGCAGATCCAAGAATATGTGGG	ACGTTGGATGTTCTAGAAAGGAGCCAAATC
rs458083	ACGTTGGATGTGTTGTTTCTTCCCCTCCTG	ACGTTGGATGTGGCTCCTTTCTAGAATCCC
rs467333	ACGTTGGATGCTTGTTATTTCTTCCCCTCC	ACGTTGGATGTTGGCTCCTTTCTAGAATCC
rs465381	ACGTTGGATGACTTGCCCATCTGTTTCCAG	ACGTTGGATGACAAGCCTCTAAGGATAGGG
rs466363	ACGTTGGATGAAGTGACCCTGAGGTGATGG	ACGTTGGATGTGAAGACAGTTACCCCGTG
rs2457099	ACGTTGGATGTCTCCTTACACTGCCAGCGT	ACGTTGGATGCACTGTATTGCTACTTGAGC
rs463901	ACGTTGGATGAGAGTGCCAAGTGCAAAGG	ACGTTGGATGTGTCTTGCGTCTGTGTATCC
rs465621	ACGTTGGATGGGAAGTCATGGAAGTGCTAG	ACGTTGGATGAAAGAGCCCTAGGCTTGGA
rs463724	ACGTTGGATGAGTGTGCCTGTCTGCCCTCA	ACGTTGGATGAAGGGCAGATGGCACACTTG
rs465242	ACGTTGGATGAGTGTGCCTGTCTGCCCTCA	ACGTTGGATGAAGGGCAGATGGCACACTTG
rs467419	ACGTTGGATGAGTCCCCAAAACGTAAGTCC	ACGTTGGATGAGTCTAATTCCCTGAGCCTC
rs456135	ACGTTGGATGAGTCTAATTCCCTGAGCCTC	ACGTTGGATGACGTAAGTCCTAATGACCGC
rs464536	ACGTTGGATGTGCTCCAGGCTTTGGTCTCT	ACGTTGGATGAATTAGACTAAGGCCATGATG
rs461898	ACGTTGGATGGGGAATACACAGCCACAGAG	ACGTTGGATGAGGTCAACGGGAACAAGGTC
rs389558	ACGTTGGATGGCAGTCTCTGACAGTTCTCTA	ACGTTGGATGTTTTCTCCCTGAAGCATGG
rs466752	ACGTTGGATGGGCCTTCTCTCCTTTAGTGC	ACGTTGGATGAGTCTCTGACAGTTCTCTAA
rs455655	ACGTTGGATGCTATTTGCACCCCATATGGC	ACGTTGGATGAACACACAGCATCAGGTTCC
rs463435	ACGTTGGATGTTTCAGCCATAGCTGGATTG	ACGTTGGATGCTCTGCTGGGAAAATGTGAC
rs2174971	ACGTTGGATGAACACAACCTTCCCCTTCGTC	ACGTTGGATGTGAATCCTTGAGGTTGAGTG
rs1979979	ACGTTGGATGTGGCTGTGAGCACCCTACTT	ACGTTGGATGCCCAAAGGAAGGGAGAATTC
rs411804	ACGTTGGATGCAGATGACAGGCGGAAAATC	ACGTTGGATGAGGCTTCCAGATGATGTCCA
rs1623885	ACGTTGGATGAATCAGCTAGGAAGAGCCTG	ACGTTGGATGTTCTCTGACCCCTCTAGGTCAG
rs1643811	ACGTTGGATGCAGGGCCCTGGTACTTTCAG	ACGTTGGATGCATGGTGGTGATTGCACCTG
rs434430	ACGTTGGATGTCCAGGAGTTCACTGTAGAG	ACGTTGGATGCACATGCATACATTATCAC
rs187538	ACGTTGGATGACATGGGGCTTGGCAAATG	ACGTTGGATGCACCTGCTCAGAAGTAGCAT
rs252067	ACGTTGGATGAGAATTGCTGTGGTGTGAGG	ACGTTGGATGTTTTCTTGGGAGCTGTCGC
rs459319	ACGTTGGATGCCATCTCTCTGACCTAGACA	ACGTTGGATGGCTCCAAGGAAAATTGGGAG
rs467289	ACGTTGGATGGGCCCTCTTGGCTTGTCTTT	ACGTTGGATGAGGCAGTGTGCCCTCTCATC



dbSNP rs#	Forward PCR primer	Reverse PCR primer
rs462644	ACGTTGGATGATGATGTGGGTGAGCCCTTG	ACGTTGGATGTAACACTCAGCACGCACCAG
rs458752	ACGTTGGATGCACCCACATCATGTGCGCTT	ACGTTGGATGCCCTTCTCTACCCAGCACTT
rs708320	ACGTTGGATGAAACCAGCCTGGCTAACATG	ACGTTGGATGACAGGTGCCTGCTATCATAC
rs457954	ACGTTGGATGAACCAGACCTTGACTGATGG	ACGTTGGATGCCTCATACAAGTAGCCAAGG
rs2411810	ACGTTGGATGGCTTAACCAGACCTTGACTG	ACGTTGGATGAGTGTAAAGGATATCCACGGC
rs3084687	ACGTTGGATGATCCCTTGAGCCAGAGATTG	ACGTTGGATGATGTCCTGTGCACACACAAG
rs69638	ACGTTGGATGTGCTCATTGCTGTCTCATC	ACGTTGGATGAGAAGAAAGGTGTGCAGTGG
rs455452	ACGTTGGATGAGTGATGATGAGCCTGCTGG	ACGTTGGATGTCAGGTTCCCTCTCTGTGTC
rs464850	ACGTTGGATGTCTCTCTGTGCTCCAGACCA	ACGTTGGATGTGGGCTGAGATTTCTGTGGG
rs431472	ACGTTGGATGAACCACTGTGGGTGTGAAGC	ACGTTGGATGAGAGACTGCATCAGGCAGGA
rs2411809	ACGTTGGATGAGCGCATAAGTGACCACCAG	ACGTTGGATGGCACTCACAGGGCATTGATG
rs2457094	ACGTTGGATGTTACTGTCACCTTGGGTCTC	ACGTTGGATGGGAAGTCTGTATAGACGCAG
rs2457095	ACGTTGGATGTTATCAAGGCCTGCGCAGTG	ACGTTGGATGACTCCTGACCTCAGGCAATC
rs2261740	ACGTTGGATGATCGTGCCACTGCACTCCAG	ACGTTGGATGTCATCTTTTGGTAGCCCCC
rs1109180	ACGTTGGATGCCAGGCCTGTATTGCACATC	ACGTTGGATGAGAATGCGTGTGCATGTGGG
rs1109179	ACGTTGGATGTGTAATGGTATGCAGACCCC	ACGTTGGATGGAGTGCCGTATTTGTCCCTTC
rs1109178	ACGTTGGATGGCAAACAACAACAGCAACAG	ACGTTGGATGAAAGTGTGGATTTGTGCAGAC
rs456909	ACGTTGGATGTAGCTGCTTCATCTGTAAAG	ACGTTGGATGGGCACCTTACCGATCTACTC
rs469124	ACGTTGGATGACTTGGACACACATAGGCTG	ACGTTGGATGTGAAATGCTCAGGGTGTGTG
rs468039	ACGTTGGATGTGAAATGCTCAGGGTGTGTG	ACGTTGGATGAGGACTTGGACACACATAGG
rs467017	ACGTTGGATGGTCTAGCTGCCACTAAACAG	ACGTTGGATGATGTGCCAAGAGGCTTTGAG
rs469290	ACGTTGGATGTGCCCTTTGTGTGCTCAGAG	ACGTTGGATGTCCCTCTGTGCTGTGTTGG
rs469090	ACGTTGGATGACTTGTCTTCAGGTGCTTGG	ACGTTGGATGGATGGTTAGTCTCCTGGTTC
rs469568	ACGTTGGATGAGCACCTCTGGCTTTCATTG	ACGTTGGATGATTCACCAGGAAATCCCAAC
rs468386	ACGTTGGATGTAATCCCAGCCCTTTGGAAG	ACGTTGGATGTATGGAGACAGGGTTTTACC
rs469349	ACGTTGGATGTTAGAGACAGAGTCTCACTC	ACGTTGGATGTTGATCCCAGGAGTTCAAGG
rs469099	ACGTTGGATGTTGGAGCTGCTCTAGTTCTC	ACGTTGGATGTGAAAACCGGAGCTCAGCTC
rs456868	ACGTTGGATGACAGAGCAGGGAGCTGCGGT	ACGTTGGATGATTCACCCCCAGCTACTGTG
rs465389	ACGTTGGATGAGGCTTTGTAGACAGCTCCC	ACGTTGGATGTGCCAGTGCTCTGAGTATGC
rs463892	ACGTTGGATGAGGCTTTGTAGACAGCTCCC	ACGTTGGATGTGCCAGTGCTCTGAGTATGC
rs468548	ACGTTGGATGACTGGAAGGGAACATGCAAG	ACGTTGGATGCCTGGATGCCCTTTATAGAC
rs654612	ACGTTGGATGACTGGAAGGGAACATGCAAG	ACGTTGGATGTGGATGCCCTTTCTAGACAC
rs468542	ACGTTGGATGGCCTCCATTTCTCTCTCAC	ACGTTGGATGTGTCTAGAAAGGGCATCCAG
rs469262	ACGTTGGATGTTCTGAGCTGAACGAAGCAG	ACGTTGGATGGGTGAGGGATCCTTTGATGC
rs708323	ACGTTGGATGCACATACTATACAGGTCACC	ACGTTGGATGGAGGGGAGAAGATGTTGTGAA
rs469089	ACGTTGGATGTTTGGAAAGTACCACCTCAGC	ACGTTGGATGAATGGAAGGAAGGATCAGCC
rs469396	ACGTTGGATGAGTGACTCCAATGAGGGAAC	ACGTTGGATGTCTCACACCACTGATCCTTC
rs468723	ACGTTGGATGTGTGGATCTTGCTGTTGGG	ACGTTGGATGTATTGGCATCGCGTATCAGG
rs467604	ACGTTGGATGACTCCTGCCATTAAACTCTC	ACGTTGGATGCTTGGCTTAACTTACAAGGG
rs338874	ACGTTGGATGCCCCACCACAGCCACTGGG	ACGTTGGATGAAGGGCCTTGCCCCACCCAA
rs338875	ACGTTGGATGTGCTGTCTTGCTCGCGTGTG	ACGTTGGATGACACTGGATATGTCAGGGTC
rs1385803	ACGTTGGATGTCACCACCATTCCAGAAGTG	ACGTTGGATGACCTTCCTTATTGCTGTGGC
rs1385804	ACGTTGGATGTCACCACCATTCCAGAAGTG	ACGTTGGATGACCTTCCTTATTGCTGTGGC
rs338876	ACGTTGGATGTTAGGGCTGGGTGGAGGAAG	ACGTTGGATGTCCAACCTCCAGTGACAGAG
rs189803	ACGTTGGATGCCTCCAGTTTCTCTCTCTG	ACGTTGGATGATCCTGGATTAGCCAGATGG
rs452215	ACGTTGGATGTAGCTCTATTCTTCCACCCC	ACGTTGGATGAGCGAGACTCCGTCTCAAAA
rs641170	ACGTTGGATGATAGCTCTATTCTTCCACCC	ACGTTGGATGAGCGAGACTCCGTCTCAAAA

dbSNP rs#	Forward PCR primer	Reverse PCR primer
rs584398	ACGTTGGATGTTCTGTGAGCTATAGAAAC	ACGTTGGATGCGAGACTCCGTCTCAAAAAAA
rs385330	ACGTTGGATGTTGCCCAACTATTGTCCTG	ACGTTGGATGGGTTTCCCAGACAGTGTTCG
rs429538	ACGTTGGATGTATTATCTGCAGACACCTGG	ACGTTGGATGATCTCATTCCCACCCTCTTC
rs371229	ACGTTGGATGTATTATCTGCAGACACCTGG	ACGTTGGATGATCTCATTCCCACCCTCTTC
rs460874	ACGTTGGATGGTCCTGCGGCTAAAAATTCC	ACGTTGGATGGGGCAGGTCAACTAGAAAAC
rs646121	ACGTTGGATGGGGCAGGTCAACTAGAAAAC	ACGTTGGATGGTCCTGCGGCTAAAAATTCC
rs468262	ACGTTGGATGGCCAGGTTTCGAAAGTTAGG	ACGTTGGATGTGGGTTGGTCATGCGGTAAC
rs467863	ACGTTGGATGTTTCGAAACCTGGCTGATGG	ACGTTGGATGTGCCACTGTCAGAAGACAAG
rs191434	ACGTTGGATGCCAGCTGAAACACTAGACAG	ACGTTGGATGAGCTGAAGAGGTCTTTCTCC
rs2054782	ACGTTGGATGAAAAAAGCAGGCCTCAGACC	ACGTTGGATGTCTGACTCTCATCTGCAGAC
rs468499	ACGTTGGATGCTCCAGGAGGGACACTACGT	ACGTTGGATGTGGCCAGCTTCTCCTCGATG
rs180287	ACGTTGGATGTTGTCTGCAGAAATTACCTAT	ACGTTGGATGGAAAAAGAAAAAAATCAG
rs338877	ACGTTGGATGCGTGGATGGAAATTTACATT	ACGTTGGATGTTCTTTGGATCAATGTTGCC
rs650665	ACGTTGGATGCCCATTCTACTCTATGATCTC	ACGTTGGATGAAAGTGCTGGGATTATAGGC
rs193419	ACGTTGGATGGCAAATCCAAAGACACAGGG	ACGTTGGATGATGTTTTTCATCACCCCAGTG
rs180288	ACGTTGGATGTGTGACCTGGTAGCTTAGAG	ACGTTGGATGTTGTAGGAGGTCAGAAGAGG
rs186834	ACGTTGGATGTAAGCTACCAGGTCACACAC	ACGTTGGATGAGTTGATAGGAGAGTCAGGC
rs189266	ACGTTGGATGTAAGCTACCAGGTCACACAC	ACGTTGGATGAGTTGATAGGAGAGTCAGGC
rs189267	ACGTTGGATGCCTCATTGTGCCCTGTTGTG	ACGTTGGATGCTCTGCCTGACTCTCCTATC
rs170937	ACGTTGGATGCCTATCAACTGTTGATGGCG	ACGTTGGATGTTCTCTATTGTGCCCTGTTG
rs463263	ACGTTGGATGTACTGGACCCCTTGCACAG	ACGTTGGATGTGCCCATGCTCATGTGTTGG
rs463262	ACGTTGGATGTGCCCATGCTCATGTGTTGG	ACGTTGGATGACCCCTTGCACAGATGCTG
rs460454	ACGTTGGATGAAGAAGGACCGTGTGAGAGA	ACGTTGGATGACATGAGCATGGGCAGGTAC
rs460455	ACGTTGGATGACATGAGCATGGGCAGGTAC	ACGTTGGATGAAGAAGGACCGTGTGAGAGA
rs460505	ACGTTGGATGACCGTGGACAGCGTCTCTGA	ACGTTGGATGTGCTCTGAGGGCAGAACAAAG
rs931316	ACGTTGGATGATGCACACACCCATGGTCAG	ACGTTGGATGCGGTTCACTCCAGCATTTCC
rs463431	ACGTTGGATGTCACCACAGCCCATGGGGA	ACGTTGGATGTTTGAAACTCACAATGTGGG
rs461542	ACGTTGGATGTGATGAAGGCCAAGAATGCT	ACGTTGGATGTGTGTCCAGAACGTCAGGTG
rs463557	ACGTTGGATGTGATGAAGGCCAAGAATGCT	ACGTTGGATGTGTGTCCAGAACGTCAGGTG
rs191453	ACGTTGGATGCATCCAACAGCTCTGTCTGC	ACGTTGGATGACCCATCTGTAGCGCATCAG
rs2271212	ACGTTGGATGAGCTTCCCCGGAGGCAACGA	ACGTTGGATGTGCAGGTCTCGGCCAAAGAC
rs462009	ACGTTGGATGCAGGCTCCTCCTCGTTGCC	ACGTTGGATGTTGGTGTCCCACGTGGTGT
rs2271211	ACGTTGGATGTCGTACCCCTGCTCTGGACG	ACGTTGGATGACTGACGCCAGGGCCGCTT
rs396474	ACGTTGGATGTGGGAGTTGGAGATGATGAG	ACGTTGGATGTTCTCAGATCCCAGTCAAG
rs428901	ACGTTGGATGTCAGTGACAGAGCGAGACTC	ACGTTGGATGGGGCTCGATAATGTAGCCAT
rs452300	ACGTTGGATGAGCACAAGCTGAAGAGGTCT	ACGTTGGATGAGGAGAGAAAGTGCACAGATC
rs670256	ACGTTGGATGTAGCTCTATTCTCCACCCC	ACGTTGGATGAGCGAGACTCCGTCTCAAAA

TABLE 12

dbSNP rs#	Extend Primer	Term Mix
rs2278221	CAAACGCTGAGGAGAAGCC	ACT
rs1650358	AAGAGACAAAAGGCCGGGC	ACT
rs1643818	TACAGGTGTGAACCACCGC	ACT
rs3733916	AGGCTGTAGTGTGACAGAC	ACG

dbSNP rs#	Extend Primer	Term Mix
rs1624933	GTCTCAAACCTCCTGACCTCA	ACT
rs1624857	AGACCAGCCTGGCCAACAT	ACT
rs1624832	GGCCAACATGGTGAAACCC	ACG
rs1624829	TGGCCAACATGGTGAAACCCT	ACT
rs2161171	TGGAATAAGAGCCCTGCAGTGG	ACT
rs1530499	CCCCTGCCCCAGCCACAGGAA	ACT
rs888764	ATGTTGTATTGGCTATATTTGTCA	ACG
rs873987	AAAACCTAAAAGAATCCACGGTA	ACG
rs4078699	GACACATGATTAACAGCAAAACAAT	ACT
rs870311	AAGGGCGTGACGGCCCC	ACT
rs1643817	GAAAGGGGAGAAAAGATTATCCC	CGT
rs1643816	AGGACCAGGAGTTTCCCATTTT	ACT
rs1650355	GAATCAATGAAGAAGAGAGCTT	ACT
rs888763	GGTCAGGAGGCAGAGGGA	ACT
rs1862212	GGGGTGAAAGGGAGCAGGG	CGT
rs1110514	CAGGCCCCAGGTGAGGAA	CGT
rs3797600	CTTTGTTGGTTAACCAAACCC	ACG
rs3797602	GCTGACAGCTCCGGACATG	ACT
rs3797603	TGTCATTCTCCTTGTGAACCCTC	ACT
rs3776819	CCATTCCATTGCACCTGCATG	ACT
rs252076	CAAAGTGCTGGGATTGCAGG	ACG
rs252075	GAGCATTTGCAGGCATGCCCTCT	ACT
rs252074	CTGGGTGGCTGCTGGGC	ACG
rs252068	GGAAGGGTTTAAGCAAGGAG	ACT
rs252069	TGAGCACCTACTATGGGCTAG	ACT
rs194040	ATTCCATATCTTCAAAGTGATTCA	ACG
rs252070	CCTGGGCTTCCCCTCCC	ACG
rs3797606	AGCCCTTGGCCTCTCTCC	ACT
rs171667	CGCCTTTTGCTTATGCAAAGA	ACG
rs187539	AACCTTCAGGAAAGTTCCCAT	ACT
rs3836834	TCAAAATATCAAACCTACCATGAAA	ACG
rs252071	ACCCTGAGACACAGGGACT	ACT
rs252072	GCTGGGTCACACTCGCGGA	ACG
rs252073	GCCGAGAAAAGTCAGGGATTCT	ACT
rs379589	CGGGAGTACTGAGCACCCAGG	CGT
rs2052472	CCCCACTGTGACTATCTCCAC	ACT
rs2052471	GTCTCCTTTGGCTGCCAAG	ACT
rs2052470	TGCCAAGGCCCTGTCCTC	ACT
rs2052469	CGCGGGGAAGTACTCGGC	ACT
rs3797608	GTCTCCTGTTCTGAGGCC	ACT
rs3797609	GGCAGAGCGGATGGCCTG	ACG
rs3822601	GTGAGGCCTGAGATGAGAACC	ACG
rs153131	TCCCCATACTCCTGTGCTC	ACG
rs751546	CCGTTCTACAGCGGTTAAGA	ACT

dbSNP rs#	Extend Primer	Term Mix
rs2279979	GGCCACCAGACAGATGTAAG	ACT
rs252060	CGTGTTTCGGCAGAGGTGA	ACT
rs3797610	CTTCTCCCTTGGGTGATGTGTT	ACT
rs194039	CCACCGTGCCGGGACATTTTTTTT	ACT
rs168773	ACTGGAGATATCACGGGAGC	CGT
rs252061	CCAGCTGGTCACAGGGCTCCC	ACG
rs187537	TCATGCTAAGTGAAATAAGCCA	ACT
rs252062	GCCACCACCGTCCACAGA	CGT
rs2431255	CTGTATATTTACCGCAATTAAAA	ACT
rs3797612	GGCATTCAATCGTCAGGGCAA	ACG
rs3797613	CCGCCGCCGGTCTCCCA	ACG
rs614114	GAACGTTCTCTCACTTTTGCC	ACT
rs252063	CTCCCGTCCTCTGAGCCTT	ACT
rs252064	GAAAACTAAGGCTCAGAGGAC	ACT
rs252065	TGGAAAAGGCGAGGCCTGGAGT	ACG
rs450502	GGAAAAGGCGAGGCCTGGAGTT	ACT
rs439252	GCCTCCCAAAGTGCTGGGATTA	ACG
rs252066	TAGCCCTCTGGAGCCCAG	ACG
rs457957	GGGCCCTCCTTAAAGCTC	ACT
rs3797614	TGGCCCTCGCTCTAATGCA	ACG
rs423552	CTCGATGTTGTAGTCATCGTC	ACG
rs398829	TGGCGTGCTCCTCTAGGA	ACG
rs416646	CTCAGCAGGTCTGATCCATC	ACT
rs187450	GGGCAGACTCCCCAGGAT	ACT
rs337807	GCAGGCCACTCGGTGGAC	ACT
rs337806	CCACCCCAGGGGTAGCCC	ACT
rs1396438	GGCAGGCAGGTGGCCTG	ACT
rs1396437	CGGCAGAAGCAGCCTCAAGA	ACG
rs2411811	ACATAATTTCCAAATTTACCCCC	CGT
rs2898813	GGTCCTGGGTGGAGGGAT	ACT
rs189256	AGCAAGGCTCTATTTGGGA	ACT
rs173072	CTTTGCTCACATCGTGGCCAAA	ACT
rs337805	GGAGCAGGAAAATTACATGACT	ACG
rs191415	GAGTCCAACGTTGTCCACAT	ACT
rs180045	ACTTGTTTCTACAATTCTCATT	ACG
rs189255	GACTCAGTCCCAGGTTTTCT	ACT
rs652766	AGAAAACCTGGGACTGAGTCT	ACT
rs466750	TCCTTAAATGCCTTGGTTGGCAAT	ACT
rs442406	TTCTGGCTGTTGGGTTTGAAC	ACT
rs662407	AAACATCTGAAATTAAGCACC	ACT
rs592971	AGCATTTCTTTGACTGCTCTTTCA	ACT
rs457187	GGAGTATCTGTTCTTGTGG	ACT
rs459490	TTGAACATAGGAATAACCCGC	ACT
rs459668	GTCTTCTTTGTGTTTTGGAGA	ACG

dbSNP rs#	Extend Primer	Term Mix
rs462646	ATTATTGACGGAGATTATTTGAC	ACT
rs458272	ATTATTTTCTGTCTGGTGTGG	ACT
rs463455	CCTCCTCATTGTCATTCTTTTC	ACT
rs675880	CTTTCATGACATTGACACAACCTAC	ACT
rs810617	CCACAGCCTCCGCCTCCC	ACT
rs464156	GGGTTTCCAGGTTAAAATGGC	ACT
rs458083	CTCCTGCTCTGCCTATCCTT	ACT
rs467333	ATTTCTTCCCCTCCTGCTCT	ACT
rs465381	GCCTCCCACAGTTCCCTTGTT	ACT
rs466363	GTGATGGCTCTGCACCAGA	ACG
rs2457099	AGCGTGTGCCAGCTCTCC	ACT
rs463901	GACACAATTCAGAGCGACTTAC	ACT
rs465621	AAGTGCTAGAAGAAAATGTAGC	ACT
rs463724	CCTTGCGCCATCCCCTAG	CGT
rs465242	TGTGCCCATCCCCCCTT	ACT
rs467419	AACGTAAGTCCTAATGACCGCCC	ACG
rs456135	CCCTCTCCTCTTCTGGGCA	ACT
rs464536	GCTTTGGTCCTCCTGAGCC	ACG
rs461898	CACAGAGCGACTCTCTCTTGTT	ACT
rs389558	CTGACAGTTCTCTAAACTCCCA	ACG
rs466752	TTCTTTTTCTCCCTGAAGCA	ACG
rs455655	CACCCCATATGGCTCATGGG	ACT
rs463435	GGGAAGGAGGTACTTAGCAG	ACG
rs2174971	GTGCCACTCTCCAGCGGCC	ACG
rs1979979	TTGCCGGCCCCCACCTC	ACG
rs411804	GAAAATCCCTGTCACCAAGTC	ACG
rs1623885	CTTGGCTGCAGCACCCCA	ACT
rs1643811	GCCCTGGTACTTTCAGCTCCCT	ACG
rs434430	GTGTGCATGTGTGTGCCTG	CGT
rs187538	TAAACGGGCCAAAAACGCCTAT	ACT
rs252067	GCGCCTACGGATGTCAGG	ACT
rs459319	CATGTTGAACAGAGAGAAACGGTC	ACG
rs467289	TCACTGAGAAATATTTTGCTCCC	ACT
rs462644	GGTGAGCCCTTGGCTGTG	ACG
rs458752	TAAAGCGCTCTTACAAATCAACA	ACT
rs708320	TGGTGAAATCCTGTCTCTACTAAA	CGT
rs457954	CACCGTTTCTTATAATGCAGCC	ACT
rs2411810	GGGGACGTTACTTCTTTTCAC	ACG
rs3084687	ATTTATATATGTGTGTGTACACAT	ACT
rs69638	CCCATTGGCTGTCCTGGAA	ACT
rs455452	CCTCAACCCAGATGCCCTC	ACG
rs464850	ACTCCTGCCTGAGTGTCTC	ACT
rs431472	GTGAAGCGGAAGGAGACTC	ACG
rs2411809	CTGCACACCCTCTGCACAG	ACG

dbSNP rs#	Extend Primer	Term Mix
rs2457094	TGGCTGGCACCACACTGCACTGC	ACT
rs2457095	TGGCTCATGCTTCTAATCCCA	ACT
rs2261740	CTGCACTCCAGCCTGGGC	ACT
rs1109180	ACATCAGTGACAGTGTAATGGTA	ACG
rs1109179	TATGCAGACCCCCTCCCC	ACT
rs1109178	AACAACAGCAACAGAAATGAAG	ACT
rs456909	CGATTCCCACGCGTGTCTG	ACG
rs469124	CCTGGCTCCATTGGTGTGAA	ACT
rs468039	CCTTCACACCAATGGAGCCAG	ACT
rs467017	CTGCCACTAAACAGATGAGAA	ACT
rs469290	ATTTCTGGGCCCAAAGTCCA	ACT
rs469090	CCAATTGTTCCAGCCACTCCC	ACT
rs469568	TGATATTGCTTGCTTGGGTCTTAG	ACT
rs468386	GGTCAAGAATTCAAGAGCAGC	ACT
rs469349	GTGCAGTGGCACGATCCTA	ACT
rs469099	GCAGGTGGAACCGCAGAC	ACT
rs456868	GGAGCTGCGGTGACTCCC	ACT
rs465389	CCCTGGCACTCGCAGACC	ACT
rs463892	AGCTCCCCCGCACCAC	ACT
rs468548	AAGGGAACATGCAAGCAAAGACTC	ACT
rs654612	TGCAAGCAAAGACTCGAATGA	ACT
rs468542	TCACTCACTTGATTCTGCCATC	ACT
rs469262	CACTGTGGGATTTCAGCAGA	ACT
rs708323	TATACAGGTCACCCATTAAAGT	ACT
rs469089	CCTCGGCCTTCCCCAGCT	ACT
rs469396	AATGAGGGAACCTGCAGTTTAAGA	ACT
rs468723	CAGACCCCATGCCTTGCC	ACT
rs467604	GAGTTTCCTCCTCTTCACAA	ACT
rs338874	CACAGCCACTGGGGAGTAG	ACT
rs338875	TTGCTCGCGTGTGCCAGCAAAT	ACG
rs1385803	AAGTGAATTCTCATGGCAGAT	ACT
rs1385804	CATTCCAGAAGTGGAATTCTCATG	ACT
rs338876	AGGAAGGTGCTCCGGCCT	ACG
rs189803	TGCTTCCCCCTTCCCCCT	CGT
rs452215	TCTATTCTTCCACCCCCATCTTT	ACT
rs641170	CTATTCTTCCACCCCCATCT	ACT
rs584398	CTCTTATATAGCTCTATTCTTCC	CGT
rs385330	AGGTGTCTGCAGATAATACATT	ACG
rs429538	CCTGGGGCACAGGACAATA	ACT
rs371229	GACAATAGTTGGGGCAAGAC	ACT
rs460874	ACAAAACATATCCTTCAAAAATACA	CGT
rs646121	GTTTTTGTCTCTGAAAGTGTCT	CGT
rs468262	CACCCAACTACTTGCTCCC	ACG
rs467863	GCTGATGGGAGGCCAATGT	ACT

dbSNP rs#	Extend Primer	Term Mix
rs191434	GTCCAGAGATCCTGCTCACT	CGT
rs2054782	CCCCCTCCATCACCTCCC	ACG
rs468499	GTGAGCCAGCAATTCTCCTA	ACT
rs180287	CAATGATCAGAACTCAGAGGTTTT	ACT
rs338877	AGAGATAAATTTCCAGTGTGAG	CGT
rs650665	AGACATCCCGGCCGGGC	ACT
rs193419	CCAAAGACACAGGGAGTAGATTA	ACT
rs180288	GAGAATATTCTTGTGGGCTTAAT	ACT
rs186834	CCAGGTCACACACACACTC	ACG
rs189266	CACACACTCCCTCTCACTGT	ACT
rs189267	TTCTGTGCATCTTTGACGCCATC	CGT
rs170937	GATGGCGTCAAAGATGCACA	ACT
rs463263	CCCCTTTGCACAGATGCTG	ACT
rs463262	GGGGAGCAGCCAGTTCCTA	ACT
rs460454	AGAGGCTGGGGACAGAGAA	ACT
rs460455	GGTACCCACCAGTCTCCTTCT	ACT
rs460505	CAGCGTCTCTGACACGGTC	ACG
rs931316	GGTCAGAGCAGACACATCCACAT	ACG
rs463431	CCCATGGGGAGCACCAAG	ACT
rs461542	TGGGAGCTCCCGGGATATTGCC	ACG
rs463557	GCTCCCGGGATATTGCCCA	ACT
rs191453	CTGGGCTGGGGCCCTGC	ACT
rs2271212	CGAGGAGGAGCCTGGCAG	ACG
rs462009	CTCCTCGTTGCCTCCGGG	ACT
rs2271211	GACGTAGCTGCCGACACCA	ACG
rs396474	CTGGTGGCCCATCTATCCTGG	ACT
rs428901	GAGCGAGACTCCGTCTCAA	CGT
rs452300	CTGAAGAGGTCTTCTCCTTCC	CGT
rs670256	TTCTTCCACCCCCATCTTTG	ACT

### Genetic Analysis

[0237] Allelotyping results from the discovery cohort are shown for cases and controls in Table 13. The allele frequency for the A2 allele is noted in the fifth and sixth columns for osteoarthritis case pools and control pools, respectively, where “AF” is allele frequency. The allele frequency for the A1 allele can be easily calculated by subtracting the A2 allele frequency from 1 (A1 AF = 1-A2 AF). For example, the SNP rs2278221 has the following case and control allele frequencies: case A1 (C) = 0.36; case A2 (T) = 0.64; control A1 (C) = 0.37; and control A2 (T) = 0.63, where the nucleotide is provided in paranthesis. Some SNPs are labeled “untyped” because of failed assays.

TABLE 13

dbSNP rs#	Position in SEQ ID NO: 1	Chromosome Position	A1/A2 Allele	F A2 Case AF	F A2 Control AF	F p- Value
rs2278221	210	178695460	C/T	0.64	0.63	0.770
rs1650358	3608	178698858	C/G			
rs1643818	3609	178698859	C/G			
rs3733916	4318	178699568	C/T			
rs1624933	5593	178700843	A/G	0.69	0.71	0.255
rs1624857	5629	178700879	C/T	0.79	0.81	0.574
rs1624832	5639	178700889	A/G	0.41	0.44	0.203
rs1624829	5640	178700890	C/T	0.89	0.93	0.044
rs2161171	8943	178704193	A/C			
rs1530499	17968	178713218	A/G	0.39	0.39	0.861
rs888764	19887	178715137	A/G			
rs873987	21034	178716284	A/G			
rs4078699	21085	178716335	C/T	0.56	0.54	0.374
rs870311	21596	178716846	A/G	0.51	0.50	0.590
rs1643817	23379	178718629	A/C	0.27	NA	NA
rs1643816	23432	178718682	A/C			
rs1650355	24007	178719257	A/C			
rs888763	26121	178721371	A/G	0.40	0.42	0.390
rs1862212	26273	178721523	A/T	0.55	0.54	0.753
rs1110514	26755	178722005	A/T	0.29	0.28	0.572
rs3797600	27411	178722661	C/T	0.56	0.57	0.738
rs3797602	27710	178722960	G/T	0.65	0.64	0.564
rs3797603	27842	178723092	C/T			
rs3776819	28379	178723629	C/T	0.46	0.46	0.850
rs252076	29603	178724853	C/T	0.46	0.48	0.519
rs252075	31232	178726482	C/G	0.35	0.36	0.859
rs252074	31504	178726754	A/G	0.35	0.34	0.816
rs252068	32583	178727833	C/G	0.47	0.48	0.656
rs252069	32794	178728044	A/G	0.28	0.27	0.626
rs194040	32840	178728090	C/T	0.31	0.32	0.665
rs252070	33044	178728294	C/T	0.58	0.57	0.573
rs3797606	33150	178728400	A/C	0.88	0.88	0.684
rs171667	33218	178728468	A/G	0.48	0.51	0.166
rs187539	33513	178728763	C/T	0.33	0.34	0.652
rs3836834	33959	178729209	- TATCA AACTAC CATGAA A			
rs252071	34486	178729736	A/G	0.30	0.31	0.666
rs252072	36289	178731539	C/T	0.49	0.50	0.677
rs252073	36570	178731820	C/T			
rs379589	38247	178733497	A/T	0.59	0.63	0.096
rs2052472	38477	178733727	A/C	0.05	0.06	0.508
rs2052471	38518	178733768	C/T	0.89	0.88	0.459
rs2052470	38529	178733779	C/T	0.83	0.80	0.125
rs2052469	38667	178733917	A/G	0.83	0.80	0.172
rs3797608	39781	178735031	C/T	0.06	0.07	0.578
rs3797609	39856	178735106	C/T	0.05	0.05	0.812
rs3822601	39927	178735177	C/T	0.08	0.08	0.802
rs153131	40506	178735756	A/G	0.76	0.77	0.944
rs751546	41869	178737119	C/G	0.93	0.92	0.585
rs2279979	42452	178737702	C/T	0.93	0.92	0.436
rs252060	44788	178740038	C/T	0.81	0.82	0.760
rs3797610	46059	178741309	A/C	0.17	0.17	0.858



dbSNP rs#	Position in SEQ ID NO: 1	Chromosome Position	A1/A2 Allele	F A2 Case AF	F A2 Control AF	F p- Value
rs194039	46846	178742096	A/G	0.41	0.47	0.035
rs168773	47712	178742962	A/T	0.35	0.38	0.266
rs252061	48796	178744046	C/T	0.21	0.19	0.508
rs187537	49441	178744691	C/G			
rs252062	49602	178744852	A/T	0.95	0.95	0.960
rs2431255	49723	178744973	A/C	0.24	0.19	0.034
rs3797612	50050	178745300	C/T	0.38	0.43	0.036
rs3797613	50171	178745421	C/T	0.21	0.21	0.941
rs614114	50477	178745727	C/T	0.50	0.53	0.387
rs252063	50818	178746068	C/T	0.57	0.55	0.313
rs252064	50833	178746083	C/T	0.52	0.52	0.806
rs252065	50881	178746131	A/G	0.22	0.22	0.857
rs450502	50882	178746132	A/G			
rs439252	51386	178746636	C/T			
rs252066	51534	178746784	C/T	0.19	0.18	0.618
rs457957	52317	178747567	A/G	0.67	0.70	0.172
rs3797614	52368	178747618	C/T			
rs423552	52970	178748220	A/G	0.90	0.92	0.215
rs398829	53023	178748273	A/G			
rs416646	53356	178748606	A/G	0.56	0.57	0.650
rs187450	53882	178749132	G/T			
rs337807	54553	178749803	C/T	0.55	0.59	0.208
rs337806	55475	178750725	A/C	0.11	0.10	0.925
rs1396438	55530	178750780	A/G	0.56	0.54	0.494
rs1396437	55691	178750941	C/T			
rs2411811	55848	178751098	A/C			
rs2898813	55879	178751129	C/G			
rs189256	56316	178751566	A/G	0.19	0.19	0.988
rs173072	56911	178752161	A/C			
rs337805	57320	178752570	A/G	0.25	0.24	0.657
rs191415	57391	178752641	C/T			
rs180045	57437	178752687	C/T	0.51	0.47	0.211
rs189255	57478	178752728	C/G	0.15	0.12	0.273
rs652766	57500	178752750	C/T	0.57	0.61	0.213
rs466750	59111	178754361	G/T	0.35	0.33	0.493
rs442406	59333	178754583	A/G	0.57	0.59	0.420
rs662407	59715	178754965	A/G	0.31	0.27	0.102
rs592971	59804	178755054	A/G			
rs457187	59851	178755101	A/G	0.23	0.24	0.842
rs459490	59929	178755179	C/T	0.21	0.20	0.604
rs459668	60052	178755302	C/T	0.20	0.19	0.648
rs462646	60240	178755490	C/T	0.43	0.43	0.905
rs458272	60359	178755609	G/T	0.22	0.20	0.523
rs463455	60381	178755631	A/G	0.25	0.24	0.644
rs675880	60456	178755706	C/T	0.63	0.65	0.591
rs810617	60724	178755974	C/G			
rs464156	60875	178756125	C/T	0.34	0.34	0.892
rs458083	60968	178756218	A/G	0.80	0.82	0.499
rs467333	60978	178756228	C/G	0.11	0.12	0.369
rs465381	60998	178756248	C/T			
rs466363	61557	178756807	C/T	0.31	0.34	0.358
rs2457099	62091	178757341	C/T	0.44	0.44	0.956
rs463901	62645	178757895	C/T	0.43	0.45	0.395
rs465621	62943	178758193	A/C	0.62	0.63	0.534
rs463724	63131	178758381	A/T	0.09	0.08	0.523
rs465242	63145	178758395	G/T			
rs467419	63406	178758656	A/G	0.65	0.66	0.647

dbSNP rs#	Position in SEQ ID NO: 1	Chromosome Position	A1/A2 Allele	F A2 Case AF	F A2 Control AF	F p- Value
rs456135	63427	178758677	C/G	0.79	0.80	0.686
rs464536	63554	178758804	C/T	0.36	0.34	0.296
rs461898	63661	178758911	A/G	0.30	0.32	0.411
rs389558	64093	178759343	A/G	0.24	0.26	0.325
rs466752	64153	178759403	C/T	0.35	0.37	0.446
rs455655	64409	178759659	C/G	0.87	0.89	0.536
rs463435	64544	178759794	C/T	0.68	0.66	0.428
rs2174971	65257	178760507	C/T	0.52	0.51	0.695
rs1979979	65626	178760876	A/G	0.07	0.06	0.692
rs411804	65739	178760989	A/G	0.78	0.78	0.976
rs1623885	66392	178761642	C/T	0.82	0.80	0.492
rs1643811	66720	178761970	C/T	0.24	0.24	0.924
rs434430	69177	178764427	A/T			
rs187538	69336	178764586	G/T			
rs252067	69636	178764886	A/G	0.21	0.23	0.606
rs459319	69823	178765073	A/G	0.19	0.20	0.640
rs467289	69928	178765178	C/T	0.26	0.26	0.988
rs462644	70547	178765797	C/T	0.59	0.58	0.914
rs458752	70633	178765883	C/T	0.18	0.20	0.513
rs708320	71805	178767055	A/C			
rs457954	72181	178767431	C/G	0.71	0.73	0.327
rs2411810	72200	178767450	C/T	0.28	0.26	0.252
rs3084687	72474	178767724	-IAT	0.13	0.12	0.884
rs69638	72567	178767817	C/G	0.54	0.52	0.449
rs455452	72973	178768223	A/G	0.59	0.60	0.733
rs464850	73468	178768718	A/G	0.11	0.09	0.249
rs431472	73889	178769139	A/G	0.33	0.34	0.713
rs2411809	75730	178770980	C/T			
rs2457094	75970	178771220	A/G	0.71	0.73	0.383
rs2457095	76114	178771364	A/G	0.74	0.76	0.551
rs2261740	76342	178771592	C/T	0.35	0.36	0.702
rs1109180	76449	178771699	A/G			
rs1109179	76465	178771715	C/T			
rs1109178	76791	178772041	A/C	0.46	0.45	0.820
rs456909	78042	178773292	A/G	0.55	0.53	0.444
rs469124	80758	178776008	A/G			
rs468039	80778	178776028	C/T			
rs467017	81356	178776606	A/C	0.33	0.32	0.665
rs469290	81576	178776826	A/G	0.57	0.57	0.871
rs469090	81689	178776939	C/T	0.82	0.83	0.387
rs469568	81759	178777009	G/T	0.38	0.38	0.888
rs468386	81950	178777200	C/G			
rs469349	82562	178777812	A/C			
rs469099	83591	178778841	C/T	0.66	0.63	0.264
rs456868	83700	178778950	A/G			
rs465389	83821	178779071	C/G			
rs463892	83842	178779092	C/G			
rs468548	83923	178779173	G/T			
rs654612	83929	178779179	A/C			
rs468542	84021	178779271	C/G			
rs469262	84175	178779425	C/T	0.45	0.47	0.405
rs708323	84417	178779667	A/G	0.73	0.69	0.138
rs469089	84747	178779997	C/G			
rs469396	85746	178780996	C/G	0.38	0.37	0.817
rs468723	86129	178781379	C/T	0.37	0.38	0.754
rs467604	86335	178781585	A/G	0.34	0.32	0.504
rs338874	87315	178782565	C/G	0.43	0.44	0.879

dbSNP rs#	Position in SEQ ID NO: 1	Chromosome Position	A1/A2 Allele	F A2 Case AF	F A2 Control AF	F p- Value
rs338875	87648	178782898	A/G	0.48	0.50	0.289
rs1385803	87764	178783014	A/C			
rs1385804	87770	178783020	C/G			
rs338876	88221	178783471	C/T	0.39	0.39	0.889
rs189803	90474	178785724	A/C			
rs452215	91148	178786398	G/T			
rs641170	91150	178786400	G/T			
rs584398	91160	178786410	G/T			
rs385330	91733	178786983	C/T			
rs429538	91772	178787022	A/C			
rs371229	91785	178787035	C/T			
rs460874	93140	178788390	A/T	0.74	0.71	0.351
rs646121	93148	178788398	A/T	0.93	0.94	0.687
rs468262	96080	178791330	A/G			
rs467863	96157	178791407	C/G			
rs191434	96313	178791563	A/C			
rs2054782	96759	178792009	C/T	0.44	0.42	0.353
rs468499	97026	178792276	A/C			
rs180287	97320	178792570	C/G			
rs338877	97732	178792982	A/T	0.04	0.04	0.863
rs650665	98713	178793963	C/G			
rs193419	99707	178794957	A/C			
rs180288	99959	178795209	C/G			
rs186834	100009	178795259	A/G			
rs189266	100020	178795270	C/G			
rs189267	100065	178795315	A/C			
rs170937	100086	178795336	C/G			
rs463263	101270	178796520	C/G			
rs463262	101276	178796526	G/T			
rs460454	101371	178796621	C/T			
rs460455	101376	178796626	C/G			
rs460505	101439	178796689	C/T			
rs931316	101820	178797070	C/T			
rs463431	102392	178797642	C/G			
rs461542	102602	178797852	A/G			
rs463557	102604	178797854	A/C			
rs191453	102896	178798146	C/T	0.11	0.14	0.123
rs2271212	189104	178884354	C/T	0.65	0.57	0.003
rs462009	189134	178884384	C/T			
rs2271211	189205	178884455	A/G			
rs396474	Not mapped	Not mapped	A/C			
rs428901	Not mapped	Not mapped	A/T	0.64	0.72	0.015
rs452300	Not mapped	Not mapped	G/T			
rs670256	Not mapped	Not mapped	G/T			

[0238] The *ADAMTS2* proximal SNPs were also allelotyped in the replication cohorts using the methods described herein and the primers provided in Tables 11 and 12. The replication allelotyping results for replication cohort #1 and replication cohort #2 are provided in Tables 14 and 15, respectively.

TABLE 14

dbSNP rs#	Position in SEQ ID NO: 1	Chromosome Position	A1/A2 Allele	F A2 Case AF	F A2 Control AF	F p- Value
rs2278221	210	178695460	C/T	0.64	0.62	0.624
rs1650358	3608	178698858	C/G			
rs1643818	3609	178698859	C/G			
rs3733916	4318	178699568	C/T			
rs1624933	5593	178700843	A/G	0.65	0.69	0.322
rs1624857	5629	178700879	C/T	0.81	untyped	NA
rs1624832	5639	178700889	A/G	0.38	0.42	0.265
rs1624829	5640	178700890	C/T	0.87	untyped	NA
rs2161171	8943	178704193	A/C			
rs1530499	17968	178713218	A/G	0.39	0.40	0.765
rs888764	19887	178715137	A/G			
rs873987	21034	178716284	A/G			
rs4078699	21085	178716335	C/T	0.55	0.54	0.733
rs870311	21596	178716846	A/G	0.50	0.50	0.828
rs1643817	23379	178718629	A/C	0.27	untyped	
rs1643816	23432	178718682	A/C			
rs1650355	24007	178719257	A/C			
rs888763	26121	178721371	A/G	0.40	0.40	0.816
rs1862212	26273	178721523	A/T	0.55	0.55	0.936
rs1110514	26755	178722005	A/T	0.29	0.29	0.997
rs3797600	27411	178722661	C/T	0.57	0.58	0.604
rs3797602	27710	178722960	G/T	0.64	0.63	0.879
rs3797603	27842	178723092	C/T			
rs3776819	28379	178723629	C/T	0.47	0.46	0.889
rs252076	29603	178724853	C/T	0.46	0.49	0.410
rs252075	31232	178726482	C/G	0.35	0.37	0.572
rs252074	31504	178726754	A/G	0.35	0.35	0.914
rs252068	32583	178727833	C/G	0.48	0.48	0.853
rs252069	32794	178728044	A/G	0.29	0.28	0.765
rs194040	32840	178728090	C/T	0.31	0.33	0.450
rs252070	33044	178728294	C/T	0.57	0.58	0.609
rs3797606	33150	178728400	A/C	0.87	0.91	0.119
rs171667	33218	178728468	A/G	0.45	0.50	0.125
rs187539	33513	178728763	C/T	0.33	0.34	0.709
rs3836834	33959	178729209	- /TATCA AACTAC CATGAA A			
rs252071	34486	178729736	A/G	0.30	0.32	0.566
rs252072	36289	178731539	C/T	0.48	0.51	0.400
rs252073	36570	178731820	C/T			
rs379589	38247	178733497	A/T	0.59	0.65	0.035
rs2052472	38477	178733727	A/C	0.04	0.06	0.493
rs2052471	38518	178733768	C/T	0.87	0.88	0.697
rs2052470	38529	178733779	C/T	0.84	0.78	0.036
rs2052469	38667	178733917	A/G	0.84	0.79	0.086
rs3797608	39781	178735031	C/T	0.06	0.07	0.530
rs3797609	39856	178735106	C/T	0.04	0.05	0.841
rs3822601	39927	178735177	C/T	0.08	0.08	0.904
rs153131	40506	178735756	A/G	0.77	0.77	0.964
rs751546	41869	178737119	C/G	0.94	0.92	0.265
rs2279979	42452	178737702	C/T	0.94	0.92	0.238
rs252060	44788	178740038	C/T	0.82	0.80	0.553
rs3797610	46059	178741309	A/C	0.16	0.18	0.459

dbSNP rs#	Position in SEQ ID NO: 1	Chromosome Position	A1/A2 Allele	F A2 Case AF	F A2 Control AF	F p- Value
rs194039	46846	178742096	A/G	0.43	0.45	0.589
rs168773	47712	178742962	A/T	0.34	0.35	0.845
rs252061	48796	178744046	C/T	0.23	0.22	0.884
rs187537	49441	178744691	C/G			
rs252062	49602	178744852	A/T	0.98	0.96	0.310
rs2431255	49723	178744973	A/C	0.24	0.19	0.108
rs3797612	50050	178745300	C/T	0.42	0.46	0.254
rs3797613	50171	178745421	C/T	0.19	0.21	0.576
rs614114	50477	178745727	C/T	0.52	0.54	0.717
rs252063	50818	178746068	C/T	0.55	0.57	0.537
rs252064	50833	178746083	C/T	0.52	0.50	0.609
rs252065	50881	178746131	A/G	0.21	0.25	0.234
rs450502	50882	178746132	A/G			
rs439252	51386	178746636	C/T			
rs252066	51534	178746784	C/T	0.20	0.20	0.883
rs457957	52317	178747567	A/G	0.66	0.71	0.162
rs3797614	52368	178747618	C/T			
rs423552	52970	178748220	A/G	0.90	0.92	0.380
rs398829	53023	178748273	A/G			
rs416646	53356	178748606	A/G	0.58	0.59	0.915
rs187450	53882	178749132	G/T			
rs337807	54553	178749803	C/T	0.60	NA	NA
rs337806	55475	178750725	A/C	0.10	0.10	0.997
rs1396438	55530	178750780	A/G	0.52	0.57	0.188
rs1396437	55691	178750941	C/T			
rs2411811	55848	178751098	A/C			
rs2898813	55879	178751129	C/G			
rs189256	56316	178751566	A/G	0.21	0.20	0.852
rs173072	56911	178752161	A/C			
rs337805	57320	178752570	A/G	0.24	0.24	0.950
rs191415	57391	178752641	C/T			
rs180045	57437	178752687	C/T	0.47	0.46	0.918
rs189255	57478	178752728	C/G	0.14	0.13	0.764
rs652766	57500	178752750	C/T	0.59	0.61	0.570
rs466750	59111	178754361	G/T	0.38	0.37	0.606
rs442406	59333	178754583	A/G	0.56	0.57	0.882
rs662407	59715	178754965	A/G	0.32	0.27	0.134
rs592971	59804	178755054	A/G			
rs457187	59851	178755101	A/G	0.23	0.25	0.451
rs459490	59929	178755179	C/T	0.22	0.21	0.671
rs459668	60052	178755302	C/T	0.20	0.19	0.712
rs462646	60240	178755490	C/T	0.42	0.44	0.439
rs458272	60359	178755609	G/T	0.21	0.21	0.755
rs463455	60381	178755631	A/G	0.25	0.25	0.783
rs675880	60456	178755706	C/T	0.62	0.63	0.741
rs810617	60724	178755974	C/G			
rs464156	60875	178756125	C/T	0.32	0.34	0.541
rs458083	60968	178756218	A/G	0.80	0.82	0.499
rs467333	60978	178756228	C/G	0.10	0.13	0.243
rs465381	60998	178756248	C/T			
rs466363	61557	178756807	C/T	0.31	0.34	0.494
rs2457099	62091	178757341	C/T	0.45	0.45	0.997
rs463901	62645	178757895	C/T	0.46	0.46	0.852
rs465621	62943	178758193	A/C	0.64	0.63	0.853
rs463724	63131	178758381	A/T	0.09	0.08	0.737
rs465242	63145	178758395	G/T			
rs467419	63406	178758656	A/G	0.64	0.65	0.694

dbSNP rs#	Position in SEQ ID NO: 1	Chromosome Position	A1/A2 Allele	F A2 Case AF	F A2 Control AF	F p- Value
rs456135	63427	178758677	C/G	0.79	0.76	0.339
rs464536	63554	178758804	C/T	0.36	0.34	0.553
rs461898	63661	178758911	A/G	0.31	0.33	0.727
rs389558	64093	178759343	A/G	0.27	0.28	0.762
rs466752	64153	178759403	C/T	0.34	0.38	0.223
rs455655	64409	178759659	C/G	0.87	untyped	NA
rs463435	64544	178759794	C/T	0.65	0.65	0.973
rs2174971	65257	178760507	C/T	0.49	0.51	0.476
rs1979979	65626	178760876	A/G	0.08	0.07	0.579
rs411804	65739	178760989	A/G	0.77	0.79	0.420
rs1623885	66392	178761642	C/T	0.81	0.78	0.451
rs1643811	66720	178761970	C/T	0.26	0.25	0.715
rs434430	69177	178764427	A/T			
rs187538	69336	178764586	G/T			
rs252067	69636	178764886	A/G	0.22	0.22	0.978
rs459319	69823	178765073	A/G	0.19	0.22	0.245
rs467289	69928	178765178	C/T	0.26	0.29	0.377
rs462644	70547	178765797	C/T	0.58	0.56	0.637
rs458752	70633	178765883	C/T	0.18	0.23	0.129
rs708320	71805	178767055	A/C			
rs457954	72181	178767431	C/G	0.69	0.73	0.143
rs2411810	72200	178767450	C/T	0.28	0.23	0.083
rs3084687	72474	178767724	-IAT	0.12	0.13	0.767
rs69638	72567	178767817	C/G	0.53	0.49	0.157
rs455452	72973	178768223	A/G	0.58	0.61	0.313
rs464850	73468	178768718	A/G	0.13	0.10	0.171
rs431472	73889	178769139	A/G	0.32	0.39	0.048
rs2411809	75730	178770980	C/T			
rs2457094	75970	178771220	A/G	0.70	0.75	0.157
rs2457095	76114	178771364	A/G	0.74	0.75	0.707
rs2261740	76342	178771592	C/T	0.34	untyped	NA
rs1109180	76449	178771699	A/G			
rs1109179	76465	178771715	C/T			
rs1109178	76791	178772041	A/C	0.47	0.48	0.715
rs456909	78042	178773292	A/G	0.56	0.54	0.537
rs469124	80758	178776008	A/G			
rs468039	80778	178776028	C/T			
rs467017	81356	178776606	A/C	0.33	0.31	0.480
rs469290	81576	178776826	A/G	0.63	0.66	0.427
rs469090	81689	178776939	C/T	0.80	0.83	0.300
rs469568	81759	178777009	G/T	0.39	0.43	0.234
rs468386	81950	178777200	C/G			
rs469349	82562	178777812	A/C			
rs469099	83591	178778841	C/T	0.66	0.60	0.066
rs456868	83700	178778950	A/G			
rs465389	83821	178779071	C/G			
rs463892	83842	178779092	C/G			
rs468548	83923	178779173	G/T			
rs654612	83929	178779179	A/C			
rs468542	84021	178779271	C/G			
rs469262	84175	178779425	C/T	0.46	0.50	0.232
rs708323	84417	178779667	A/G	0.72	0.66	0.071
rs469089	84747	178779997	C/G			
rs469396	85746	178780996	C/G	0.37	0.35	0.522
rs468723	86129	178781379	C/T	0.39	0.41	0.495
rs467604	86335	178781585	A/G	0.33	0.30	0.303
rs338874	87315	178782565	C/G	0.44	0.46	0.628

dbSNP rs#	Position in SEQ ID NO: 1	Chromosome Position	A1/A2 Allele	F A2 Case AF	F A2 Control AF	F p- Value
rs338875	87648	178782898	A/G	0.49	0.54	0.106
rs1385803	87764	178783014	A/C			
rs1385804	87770	178783020	C/G			
rs338876	88221	178783471	C/T	0.38	0.36	0.609
rs189803	90474	178785724	A/C			
rs452215	91148	178786398	G/T			
rs641170	91150	178786400	G/T			
rs584398	91160	178786410	G/T			
rs385330	91733	178786983	C/T			
rs429538	91772	178787022	A/C			
rs371229	91785	178787035	C/T			
rs460874	93140	178788390	A/T	0.74	0.69	0.118
rs646121	93148	178788398	A/T	0.93	0.95	0.477
rs468262	96080	178791330	A/G			
rs467863	96157	178791407	C/G			
rs191434	96313	178791563	A/C			
rs2054782	96759	178792009	C/T	0.45	0.42	0.514
rs468499	97026	178792276	A/C			
rs180287	97320	178792570	C/G			
rs338877	97732	178792982	A/T	0.04	0.04	0.781
rs650665	98713	178793963	C/G			
rs193419	99707	178794957	A/C			
rs180288	99959	178795209	C/G			
rs186834	100009	178795259	A/G			
rs189266	100020	178795270	C/G			
rs189267	100065	178795315	A/C			
rs170937	100086	178795336	C/G			
rs463263	101270	178796520	C/G			
rs463262	101276	178796526	G/T			
rs460454	101371	178796621	C/T			
rs460455	101376	178796626	C/G			
rs460505	101439	178796689	C/T			
rs931316	101820	178797070	C/T			
rs463431	102392	178797642	C/G			
rs461542	102602	178797852	A/G			
rs463557	102604	178797854	A/C			
rs191453	102896	178798146	C/T	0.15	0.19	0.139
rs2271212	189104	178884354	C/T	0.64	0.58	0.072
rs462009	189134	178884384	C/T			
rs2271211	189205	178884455	A/G			
rs396474	Not mapped	Not mapped	A/C			
rs428901	Not mapped	Not mapped	A/T	0.66	untyped	NA
rs452300	Not mapped	Not mapped	G/T			
rs670256	Not mapped	Not mapped	G/T			

TABLE 15

dbSNP rs#	Position in SEQ ID NO: 1	Chromosome Position	A1/A2 Allele	F A2 Case AF	F A2 Control AF	F p- Value
rs2278221	210	178695460	C/T	0.64	0.64	0.837
rs1650358	3608	178698858	C/G			
rs1643818	3609	178698859	C/G			
rs3733916	4318	178699568	C/T			
rs1624933	5593	178700843	A/G	0.73	0.75	0.447
rs1624857	5629	178700879	C/T	0.78	0.81	0.289

dbSNP rs#	Position in SEQ ID NO: 1	Chromosome Position	A1/A2 Allele	F A2 Case AF	F A2 Control AF	F p- Value
rs1624832	5639	178700889	A/G	0.44	0.47	0.423
rs1624829	5640	178700890	C/T	0.90	0.93	0.294
rs2161171	8943	178704193	A/C			
rs1530499	17968	178713218	A/G	0.39	0.36	0.499
rs888764	19887	178715137	A/G			
rs873987	21034	178716284	A/G			
rs4078699	21085	178716335	C/T	0.57	0.54	0.316
rs870311	21596	178716846	A/G	0.52	0.50	0.579
rs1643817	23379	178718629	A/C			
rs1643816	23432	178718682	A/C			
rs1650355	24007	178719257	A/C			
rs888763	26121	178721371	A/G	0.40	0.44	0.264
rs1862212	26273	178721523	A/T	0.56	0.53	0.529
rs1110514	26755	178722005	A/T	0.30	0.27	0.381
rs3797600	27411	178722661	C/T	0.55	0.54	0.840
rs3797602	27710	178722960	G/T	0.68	0.65	0.534
rs3797603	27842	178723092	C/T			
rs3776819	28379	178723629	C/T	0.45	0.47	0.662
rs252076	29603	178724853	C/T	0.46	0.46	0.986
rs252075	31232	178726482	C/G	0.36	0.34	0.666
rs252074	31504	178726754	A/G	0.35	0.33	0.604
rs252068	32583	178727833	C/G	0.47	0.48	0.648
rs252069	32794	178728044	A/G	0.27	0.26	0.640
rs194040	32840	178728090	C/T	0.31	0.30	0.734
rs252070	33044	178728294	C/T	0.61	0.55	0.157
rs3797606	33150	178728400	A/C	0.91	0.83	0.005
rs171667	33218	178728468	A/G	0.51	0.52	0.674
rs187539	33513	178728763	C/T	0.32	0.33	0.836
rs3836834	33959	178729209	- TATCA AACTAC CATGAA A			
rs252071	34486	178729736	A/G	0.30	0.30	0.942
rs252072	36289	178731539	C/T	0.50	0.49	0.684
rs252073	36570	178731820	C/T			
rs379589	38247	178733497	A/T	0.60	0.61	0.981
rs2052472	38477	178733727	A/C	0.06	0.06	0.856
rs2052471	38518	178733768	C/T	0.91	0.86	0.079
rs2052470	38529	178733779	C/T	0.82	0.83	0.828
rs2052469	38667	178733917	A/G	0.82	0.82	0.983
rs3797608	39781	178735031	C/T	0.06	0.06	0.969
rs3797609	39856	178735106	C/T	0.05	0.05	0.879
rs3822601	39927	178735177	C/T	0.07	0.08	0.838
rs153131	40506	178735756	A/G	0.76	0.76	0.981
rs751546	41869	178737119	C/G	0.91	0.92	0.526
rs2279979	42452	178737702	C/T	0.92	0.92	0.906
rs252060	44788	178740038	C/T	0.81	0.85	0.157
rs3797610	46059	178741309	A/C	0.18	0.16	0.593
rs194039	46846	178742096	A/G	0.39	0.49	0.005
rs168773	47712	178742962	A/T	0.37	0.43	0.098
rs252061	48796	178744046	C/T	0.19	0.15	0.164
rs187537	49441	178744691	C/G			
rs252062	49602	178744852	A/T	0.93	0.95	0.290
rs2431255	49723	178744973	A/C	0.23	0.19	0.201
rs3797612	50050	178745300	C/T	0.32	0.38	0.102
rs3797613	50171	178745421	C/T	0.23	NA	



dbSNP rs#	Position in SEQ ID NO: 1	Chromosome Position	A1/A2 Allele	F A2 Case AF	F A2 Control AF	F p- Value
rs614114	50477	178745727	C/T	0.48	0.51	0.423
rs252063	50818	178746068	C/T	0.60	0.51	0.011
rs252064	50833	178746083	C/T	0.51	0.56	0.265
rs252065	50881	178746131	A/G	0.22	0.18	0.175
rs450502	50882	178746132	A/G			
rs439252	51386	178746636	C/T			
rs252066	51534	178746784	C/T	0.18	0.16	0.451
rs457957	52317	178747567	A/G	0.67	0.68	0.728
rs3797614	52368	178747618	C/T			
rs423552	52970	178748220	A/G	0.89	0.91	0.398
rs398829	53023	178748273	A/G			
rs416646	53356	178748606	A/G	0.54	0.55	0.643
rs187450	53882	178749132	G/T			
rs337807	54553	178749803	C/T	0.49	0.59	0.009
rs337806	55475	178750725	A/C	0.11	0.10	0.889
rs1396438	55530	178750780	A/G	0.61	0.50	0.007
rs1396437	55691	178750941	C/T			
rs2411811	55848	178751098	A/C			
rs2898813	55879	178751129	C/G			
rs189256	56316	178751566	A/G	0.17	0.17	0.923
rs173072	56911	178752161	A/C			
rs337805	57320	178752570	A/G	0.27	0.25	0.582
rs191415	57391	178752641	C/T			
rs180045	57437	178752687	C/T	0.56	0.48	0.115
rs189255	57478	178752728	C/G	0.16	0.12	0.168
rs652766	57500	178752750	C/T	0.55	0.61	0.231
rs466750	59111	178754361	G/T	0.31	0.28	0.473
rs442406	59333	178754583	A/G	0.58	0.63	0.209
rs662407	59715	178754965	A/G	0.30	0.28	0.449
rs592971	59804	178755054	A/G			
rs457187	59851	178755101	A/G	0.23	0.21	0.402
rs459490	59929	178755179	C/T	0.20	0.19	0.708
rs459668	60052	178755302	C/T	0.21	0.20	0.821
rs462646	60240	178755490	C/T	0.44	0.41	0.460
rs458272	60359	178755609	G/T	0.22	0.20	0.524
rs463455	60381	178755631	A/G	0.23	0.22	0.629
rs675880	60456	178755706	C/T	0.65	0.67	0.564
rs810617	60724	178755974	C/G			
rs464156	60875	178756125	C/T	0.37	0.34	0.439
rs458083	60968	178756218	A/G			
rs467333	60978	178756228	C/G	0.11	0.11	0.902
rs465381	60998	178756248	C/T			
rs466363	61557	178756807	C/T	0.32	0.34	0.547
rs2457099	62091	178757341	C/T	0.43	0.43	0.974
rs463901	62645	178757895	C/T	0.39	0.43	0.342
rs465621	62943	178758193	A/C	0.59	0.64	0.195
rs463724	63131	178758381	A/T	0.09	0.07	0.539
rs465242	63145	178758395	G/T			
rs467419	63406	178758656	A/G	0.66	0.67	0.752
rs456135	63427	178758677	C/G	0.79	0.85	0.029
rs464536	63554	178758804	C/T	0.36	0.32	0.332
rs461898	63661	178758911	A/G	0.28	0.31	0.423
rs389558	64093	178759343	A/G	0.20	0.23	0.311
rs466752	64153	178759403	C/T	0.36	0.35	0.781
rs455655	64409	178759659	C/G	NA	0.72	NA
rs463435	64544	178759794	C/T	0.72	0.68	0.230
rs2174971	65257	178760507	C/T	0.56	0.51	0.142

dbSNP rs#	Position in SEQ ID NO: 1	Chromosome Position	A1/A2 Allele	F A2 Case AF	F A2 Control AF	F p- Value
rs1979979	65626	178760876	A/G	0.05	0.05	0.993
rs411804	65739	178760989	A/G	0.80	0.77	0.343
rs1623885	66392	178761642	C/T	0.84	0.84	0.819
rs1643811	66720	178761970	C/T	0.22	0.23	0.847
rs434430	69177	178764427	A/T			
rs187538	69336	178764586	G/T			
rs252067	69636	178764886	A/G	0.21	0.24	0.369
rs459319	69823	178765073	A/G	0.18	0.15	0.353
rs467289	69928	178765178	C/T	0.27	0.22	0.179
rs462644	70547	178765797	C/T	0.60	0.61	0.609
rs458752	70633	178765883	C/T	0.18	0.15	0.271
rs708320	71805	178767055	A/C			
rs457954	72181	178767431	C/G	0.72	0.72	0.882
rs2411810	72200	178767450	C/T	0.29	0.30	0.630
rs3084687	72474	178767724	-IAT	0.13	0.11	0.509
rs69638	72567	178767817	C/G	0.54	0.57	0.440
rs455452	72973	178768223	A/G	0.60	0.58	0.499
rs464850	73468	178768718	A/G	0.10	0.09	0.839
rs431472	73889	178769139	A/G	0.35	0.27	0.025
rs2411809	75730	178770980	C/T			
rs2457094	75970	178771220	A/G	0.71	0.70	0.792
rs2457095	76114	178771364	A/G	0.75	0.76	0.602
rs2261740	76342	178771592	C/T	0.36	0.36	0.924
rs1109180	76449	178771699	A/G			
rs1109179	76465	178771715	C/T			
rs1109178	76791	178772041	A/C	0.45	0.42	0.420
rs456909	78042	178773292	A/G	0.53	0.51	0.598
rs469124	80758	178776008	A/G			
rs468039	80778	178776028	C/T			
rs467017	81356	178776606	A/C	0.34	0.35	0.762
rs469290	81576	178776826	A/G	0.49	0.44	0.223
rs469090	81689	178776939	C/T	0.83	0.84	0.883
rs469568	81759	178777009	G/T	0.36	0.30	0.115
rs468386	81950	178777200	C/G			
rs469349	82562	178777812	A/C			
rs469099	83591	178778841	C/T	0.65	0.67	0.560
rs456868	83700	178778950	A/G			
rs465389	83821	178779071	C/G			
rs463892	83842	178779092	C/G			
rs468548	83923	178779173	G/T			
rs654612	83929	178779179	A/C			
rs468542	84021	178779271	C/G			
rs469262	84175	178779425	C/T	0.45	0.43	0.762
rs708323	84417	178779667	A/G	0.74	0.74	0.899
rs469089	84747	178779997	C/G			
rs469396	85746	178780996	C/G	0.39	0.42	0.569
rs468723	86129	178781379	C/T	0.36	0.34	0.573
rs467604	86335	178781585	A/G	0.35	0.36	0.763
rs338874	87315	178782565	C/G	0.42	0.40	0.564
rs338875	87648	178782898	A/G	0.46	0.45	0.701
rs1385803	87764	178783014	A/C			
rs1385804	87770	178783020	C/G			
rs338876	88221	178783471	C/T	0.41	0.44	0.580
rs189803	90474	178785724	A/C			
rs452215	91148	178786398	G/T			
rs641170	91150	178786400	G/T			
rs584398	91160	178786410	G/T			

dbSNP rs#	Position in SEQ ID NO: 1	Chromosome Position	A1/A2 Allele	F A2 Case AF	F A2 Control AF	F p- Value
rs385330	91733	178786983	C/T			
rs429538	91772	178787022	A/C			
rs371229	91785	178787035	C/T			
rs460874	93140	178788390	A/T	0.73	0.75	0.550
rs646121	93148	178788398	A/T	0.93	0.92	0.697
rs468262	96080	178791330	A/G			
rs467863	96157	178791407	C/G			
rs191434	96313	178791563	A/C			
rs2054782	96759	178792009	C/T	0.43	0.40	0.473
rs468499	97026	178792276	A/C			
rs180287	97320	178792570	C/G			
rs338877	97732	178792982	A/T	0.04	0.04	0.928
rs650665	98713	178793963	C/G			
rs193419	99707	178794957	A/C			
rs180288	99959	178795209	C/G			
rs186834	100009	178795259	A/G			
rs189266	100020	178795270	C/G			
rs189267	100065	178795315	A/C			
rs170937	100086	178795336	C/G			
rs463263	101270	178796520	C/G			
rs463262	101276	178796526	G/T			
rs460454	101371	178796621	C/T			
rs460455	101376	178796626	C/G			
rs460505	101439	178796689	C/T			
rs931316	101820	178797070	C/T			
rs463431	102392	178797642	C/G			
rs461542	102602	178797852	A/G			
rs463557	102604	178797854	A/C			
rs191453	102896	178798146	C/T	0.06	0.06	0.929
rs2271212	189104	178884354	C/T	0.66	0.56	<b>0.012</b>
rs462009	189134	178884384	C/T			
rs2271211	189205	178884455	A/G			
rs396474	Not mapped	Not mapped	A/C			
rs428901	Not mapped	Not mapped	A/T	0.61	0.72	<b>0.002</b>
rs452300	Not mapped	Not mapped	G/T			
rs670256	Not mapped	Not mapped	G/T			

[0239] Allelotyping results were considered particularly significant with a calculated p-value of less than or equal to 0.05 for allelotype results. These values are indicated in bold. The allelotyping p-values were plotted in Figure 1 for the discovery cohort. The position of each SNP on the chromosome is presented on the x-axis. The y-axis gives the negative logarithm (base 10) of the p-value comparing the estimated allele in the case group to that of the control group. The minor allele frequency of the control group for each SNP designated by an X or other symbol on the graphs in Figure 1 can be determined by consulting Table 13. For example, the left-most X on the left graph is at position 178695460. By proceeding down the Table from top to bottom and across the graphs from left to right the allele frequency associated with each symbol shown can be determined.

[0240] To aid the interpretation, multiple lines have been added to the graph. The broken horizontal lines are drawn at two common significance levels, 0.05 and 0.01. The vertical broken lines are drawn

every 20kb to assist in the interpretation of distances between SNPs. Two other lines are drawn to expose linear trends in the association of SNPs to the disease. The light gray line (or generally bottom-most curve) is a nonlinear smoother through the data points on the graph using a local polynomial regression method (W.S. Cleveland, E. Grosse and W.M. Shyu (1992) Local regression models. Chapter 8 of Statistical Models in S eds J.M. Chambers and T.J. Hastie, Wadsworth & Brooks/Cole.). The black line provides a local test for excess statistical significance to identify regions of association. This was created by use of a 10kb sliding window with 1kb step sizes. Within each window, a chi-square goodness of fit test was applied to compare the proportion of SNPs that were significant at a test wise level of 0.01, to the proportion that would be expected by chance alone (0.05 for the methods used here). Resulting p-values that were less than  $10^{-8}$  were truncated at that value.

[0241] Finally, the exons and introns of the genes in the covered region are plotted below each graph at the appropriate chromosomal positions. The gene boundary is indicated by the broken horizontal line. The exon positions are shown as thick, unbroken bars. An arrow is placed at the 3' end of each gene to show the direction of transcription.

#### Example 5

##### Effect of ADAMTS2 Polypeptides on Biosynthesis of Type II Collagen in Patients with OA

[0242] To investigate the effect of ADAMTS2 polypeptides on Type II collagen biosynthesis and processing, human articular cartilage from OA patients undergoing joint replacement is harvested, dissected and maintained as described by Nelson *et al.* (1998) *supra*. Type II procollagen levels in osteoarthritic patients and autopsy controls is determined by radioimmunoassay (RIA) as previously described. Allelic variations (*e.g.*, rs398829) are determined for the OA patients and controls by genotyping (See Examples 1 and 2). As type II procollagen is processed by ADAMTS2, increased levels of Type II procollagen in individuals with the allelic variation associated with OA demonstrates that this variation leads to reduced procollagen processing activity and ultimately to OA.

#### Example 6

##### Effect of ADAMTS2 Polypeptides on Type II Collagen Processing Activity

[0243] To investigate the effect of ADAMTS2 polypeptide variants on ADAMTS2 collagen processing activity, recombinant polypeptides encompassing the ADAMTS2 variation of SEQ ID NO: 2 at position 733 and a wild-type ADAMTS2 polypeptide are expressed in cell lines such as chondrocytes. Since the allelic variation of ADAMTS2 at position 733 of SEQ ID: NO: 2 will prevent the conversion of the ADAMTS2 pro-enzyme to the catalytically active enzyme, processing of ADAMTS2 pro-enzyme is monitored by SDS-PAGE analysis followed by Western Blotting using antibodies to ADAMTS2 and

methods common to someone skilled in the art. Reduced levels of pro-enzyme cleavage are apparent by the increased levels of immunopositive protein of higher molecular weight than of the cleaved active protein.

### Example 7

#### *In Vitro* Production of Target Polypeptides

[0244] cDNA is cloned into a pIVEX 2.3-MCS vector (Roche Biochem) using a directional cloning method. A cDNA insert is prepared using PCR with forward and reverse primers having 5' restriction site tags (in frame) and 5-6 additional nucleotides in addition to 3' gene-specific portions, the latter of which is typically about twenty to about twenty-five base pairs in length. A Sal I restriction site is introduced by the forward primer and a Sma I restriction site is introduced by the reverse primer. The ends of PCR products are cut with the corresponding restriction enzymes (*i.e.*, Sal I and Sma I) and the products are gel-purified. The pIVEX 2.3-MCS vector is linearized using the same restriction enzymes, and the fragment with the correct sized fragment is isolated by gel-purification. Purified PCR product is ligated into the linearized pIVEX 2.3-MCS vector and *E. coli* cells transformed for plasmid amplification. The newly constructed expression vector is verified by restriction mapping and used for protein production.

[0245] *E. coli* lysate is reconstituted with 0.25 ml of Reconstitution Buffer, the Reaction Mix is reconstituted with 0.8 ml of Reconstitution Buffer; the Feeding Mix is reconstituted with 10.5 ml of Reconstitution Buffer; and the Energy Mix is reconstituted with 0.6 ml of Reconstitution Buffer. 0.5 ml of the Energy Mix was added to the Feeding Mix to obtain the Feeding Solution. 0.75 ml of Reaction Mix, 50  $\mu$ l of Energy Mix, and 10  $\mu$ g of the template DNA is added to the *E. coli* lysate.

[0246] Using the reaction device (Roche Biochem), 1 ml of the Reaction Solution is loaded into the reaction compartment. The reaction device is turned upside-down and 10 ml of the Feeding Solution is loaded into the feeding compartment. All lids are closed and the reaction device is loaded into the RTS500 instrument. The instrument is run at 30°C for 24 hours with a stir bar speed of 150 rpm. The pIVEX 2.3 MCS vector includes a nucleotide sequence that encodes six consecutive histidine amino acids on the C-terminal end of the target polypeptide for the purpose of protein purification. Target polypeptide is purified by contacting the contents of reaction device with resin modified with Ni<sup>2+</sup> ions. Target polypeptide is eluted from the resin with a solution containing free Ni<sup>2+</sup> ions.

Example 8

Cellular Production of Target Polypeptides

[0247] Nucleic acids are cloned into DNA plasmids having phage recombination sites and target polypeptides are expressed therefrom in a variety of host cells. Alpha phage genomic DNA contains short sequences known as attP sites, and *E. coli* genomic DNA contains unique, short sequences known as attB sites. These regions share homology, allowing for integration of phage DNA into *E. coli* via directional, site-specific recombination using the phage protein Int and the *E. coli* protein IHF. Integration produces two new att sites, L and R, which flank the inserted prophage DNA. Phage excision from *E. coli* genomic DNA can also be accomplished using these two proteins with the addition of a second phage protein, Xis. DNA vectors have been produced where the integration/excision process is modified to allow for the directional integration or excision of a target DNA fragment into a backbone vector in a rapid *in vitro* reaction (Gateway™ Technology (Invitrogen, Inc.)).

[0248] A first step is to transfer the nucleic acid insert into a shuttle vector that contains attL sites surrounding the negative selection gene, ccdB (*e.g.* pENTER vector, Invitrogen, Inc.). This transfer process is accomplished by digesting the nucleic acid from a DNA vector used for sequencing, and to ligate it into the multicloning site of the shuttle vector, which will place it between the two attL sites while removing the negative selection gene ccdB. A second method is to amplify the nucleic acid by the polymerase chain reaction (PCR) with primers containing attB sites. The amplified fragment then is integrated into the shuttle vector using Int and IHF. A third method is to utilize a topoisomerase-mediated process, in which the nucleic acid is amplified via PCR using gene-specific primers with the 5' upstream primer containing an additional CACC sequence (*e.g.*, TOPO® expression kit (Invitrogen, Inc.)). In conjunction with Topoisomerase I, the PCR amplified fragment can be cloned into the shuttle vector via the attL sites in the correct orientation.

[0249] Once the nucleic acid is transferred into the shuttle vector, it can be cloned into an expression vector having attR sites. Several vectors containing attR sites for expression of target polypeptide as a native polypeptide, N-fusion polypeptide, and C-fusion polypeptides are commercially available (*e.g.*, pDEST (Invitrogen, Inc.)), and any vector can be converted into an expression vector for receiving a nucleic acid from the shuttle vector by introducing an insert having an attR site flanked by an antibiotic resistant gene for selection using the standard methods described above. Transfer of the nucleic acid from the shuttle vector is accomplished by directional recombination using Int, IHF, and Xis (LR clonase). Then the desired sequence can be transferred to an expression vector by carrying out a one hour incubation at room temperature with Int, IHF, and Xis, a ten minute incubation at 37°C with proteinase K, transforming bacteria and allowing expression for one hour, and then plating on selective media. Generally, 90% cloning efficiency is achieved by this method. Examples of expression vectors

are pDEST 14 bacterial expression vector with att7 promoter, pDEST 15 bacterial expression vector with a T7 promoter and a N-terminal GST tag, pDEST 17 bacterial vector with a T7 promoter and a N-terminal polyhistidine affinity tag, and pDEST 12.2 mammalian expression vector with a CMV promoter and neo resistance gene. These expression vectors or others like them are transformed or transfected into cells for expression of the target polypeptide or polypeptide variants. These expression vectors are often transfected, for example, into murine-transformed adipocyte cell line 3T3-L1, (ATCC), human embryonic kidney cell line 293, and rat cardiomyocyte cell line H9C2.

[0250] Modifications may be made to the foregoing without departing from the basic aspects of the invention. Although the invention has been described in substantial detail with reference to one or more specific embodiments, those of skill in the art will recognize that changes may be made to the embodiments specifically disclosed in this application, yet these modifications and improvements are within the scope and spirit of the invention, as set forth in the claims which follow. All publications or patent documents cited in this specification are incorporated herein by reference as if each such publication or document was specifically and individually indicated to be incorporated herein by reference.

[0251] Citation of the above publications or documents is not intended as an admission that any of the foregoing is pertinent prior art, nor does it constitute any admission as to the contents or date of these publications or documents. U.S. patents and other publications referenced herein are hereby incorporated by reference.

#### Nucleotide and Amino Acid Sequence Embodiments

[0252] Table A includes information pertaining to the incident polymorphic variant associated with osteoarthritis identified herein. Public information pertaining to the polymorphism and the genomic sequence that includes the polymorphism are indicated. The genomic sequences identified in Table A may be accessed at the http address [www.ncbi.nih.gov/entrez/query.fcgi](http://www.ncbi.nih.gov/entrez/query.fcgi), for example, by using the publicly available SNP reference number (*e.g.*, rs398829). The chromosome position refers to the position of the SNP within NCBI's Genome Build 34, which may be accessed at the following http address: [www.ncbi.nlm.nih.gov/mapview/map\\_search.cgi?chr=hum\\_chr.inf&query=](http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?chr=hum_chr.inf&query=). The "Contig Position" provided in Table A corresponds to a nucleotide position set forth in the contig sequence (see "Contig Accession No."), and designates the polymorphic site corresponding to the SNP reference number. The sequence containing the polymorphisms also may be referenced by the "Nucleotide Accession No." set forth in Table A. The "Sequence Identification" corresponds to cDNA sequence that encodes associated target polypeptides (*e.g.*, ADAMTS2). The position of the SNP within the cDNA sequence is provided in the "Sequence Position" column of Table A. If the SNP falls within an exon, the corresponding amino acid position (and amino acid change, if applicable) is provided as well. Also, the

allelic variation at the polymorphic site and the allelic variant identified as associated with osteoarthritis is specified in Table A. All nucleotide and polypeptide sequences referenced and accessed by the parameters set forth in Table A are incorporated herein by reference.

Table A

RS_ID	Chrom- osome	Chrom Position	Contig Accession No. [1]	Contig Position	Nucleotide Accession No. [2]	Sequence Position	Amino Acid Position	Locus	Locus ID	A [3]	Allelic Vari- ability	OA Assoc. Allele
rs398829	5	178748273	Hs5_77500_34:3	1729878	NM_014244	coding- nonsynon	V245I	ADAMTS2	9509	R	[G/A]	G

[1] Contig Accession Number which can be found in the NCBI Database:  
http address: [www.ncbi.nih.gov/entrez/query.fcgi](http://www.ncbi.nih.gov/entrez/query.fcgi)

[2] Sequence Identification or Nucleotide Accession Number which can be found in the NCBI Database:  
http address: [www.ncbi.nih.gov/entrez/query.fcgi](http://www.ncbi.nih.gov/entrez/query.fcgi)

[3] "A" column is the sequence orientation ("F" is forward, "R" is reverse).

[0253] The following is a genomic nucleotide sequence for an *ADAMTS2* region. The following nucleotide representations are used throughout: "A" or "a" is adenosine, adenine, or adenylic acid; "C" or "c" is cytidine, cytosine, or cytidylic acid; "G" or "g" is guanosine, guanine, or guanylic acid; "T" or "t" is thymidine, thymine, or thymidylic acid; and "I" or "i" is inosine, hypoxanthine, or inosinic acid. Exons are indicated in italicized lower case type, introns are depicted in normal text lower case type, and polymorphic sites are depicted in bold upper case type. SNPs are designated by the following convention: "R" represents A or G, "M" represents A or C; "W" represents A or T; "Y" represents C or T; "S" represents C or G; "K" represents G or T; "V" represents A, C or G; "H" represents A, C, or T; "D" represents A, G, or T; "B" represents C, G, or T; and "N" represents A, G, C, or T.

#### ADAMTS2 Genomic Sequence (SEQ ID NO. 1)

>5:178695251-178884700

```

1      gggcctcggg ccagcactgc ccagcgctgg gaagacagga gaccacaccc caagtggcctt
61      tgacacaggg cgtgtccctc ttaaggcaca gaggagaagt gggcagccag ggctggggat
121     cctaggggtgg cccctctgtg ccccaccctt ccccaggcca cttacacgtg gccagtctca
181     tgggccacca caaacgctga ggagaagccr tctctatggt tcaggggtgca gctgcggacc
241     ggatggcaca tgccggtgac aggagcatag cctgggagga gacaagaggc ggctccagat
301     gctgccatag cctggccggg aagggtgagg ctggcctaac tcccaggcgc tgcttctcct
361     caaggcccca atgcctctc taccacaggc aactgcccc ggctggcaca gctcagaaga
421     caccacagca gacagctcat agcctgtgac atctggctga cagcaggccc ccagcccgtc
481     accacacaag ccccgtaggc gttctgctg taacagccac actgcagctg gggccctctg
541     cttagtccaga tgtcacctgc tatggcttgt ctgacacctg actttcctgc ctgccgttct
601     ctttcacgcc acctgcctct gcccctgcc tgccctactg agtgggccctt gaggcagacc
661     tgcttccaca cctctctctg ctagaggcat cctgcccctc gtctgcctgg tatcacaccc
721     tccaagccca cctcctctgg ggggcgttcc ttgggtctcc ccgtgtttct ctgggggtcct
781     tgagacccta gctagacctg tttccataat gccgcagaag gctgcaaagc tgttcagccg

```



841	atggggccaca	ctaccctgca	gaatatccca	gcagggggct	ggctgggtgcg	gactgggagg
901	tcagagccgc	cgtagaagat	gaggccagtg	gcctacgaca	gcccagggcc	agggggaggg
961	tggcagaagt	tcatgacgca	gagacccac	gtcaaaactga	ggaagatgga	taatgagacc
1021	tattcacttc	cccagctcct	gtctcagcag	agggaggcgg	cttggaaga	tagggaagtg
1081	cctacagggc	tatgggtgcc	acacgaggtc	tctgcttgta	cccctgggag	tcttcctaga
1141	ggcctgccaa	gcatggagcc	ctgactttca	gaacccccac	ctgcccctac	tccctgccatg
1201	ccccactgag	agaccctgag	gccacgcgtt	ccacagagaa	gggaaggagg	gccactggga
1261	ccccgggtga	gctggcggtg	tacgcgccta	aaaatagaag	catcgtgatc	taaccttgt
1321	cttttcccgt	tgtataaaac	aggaaaatct	gcattttttg	ctctgcagat	tttgagacga
1381	gcgtggagtc	tgtgtgcctg	agtggaaagt	tgccatctcc	tctttgccct	ttgtatgcca
1441	ggcttgacga	tgcactctgc	tccgtgggag	gctttccatt	catcctactc	agcattacag
1501	gagcctgcga	gatctgaaag	cgcactctta	tgtttttctt	tgtttttctt	cctcttattt
1561	tactttgtgg	ctgttttctc	tctccggaat	cttatttagat	agagtgtgcc	tccctcgatgg
1621	agggctctgtc	tcctaacttt	tctctcatgt	taatcttttt	gtcttttgct	cccattggga
1681	gatggccttg	actttttatt	tcagctttcg	attttttttt	tttttttttg	agatggagtc
1741	tcactctgcc	gtccaggtcg	gagtgcagtg	gtgtgatctc	ggctcactgc	aacctccgcc
1801	tcctgggttc	agtccattct	actgectcag	cctcctgagt	agctgggact	acaggcaccc
1861	ggcccatgac	tacgctaatt	tttgtatttt	tactagagac	gggattttcac	catgttagcc
1921	aggatggctc	ccatctcctg	acctcgtgac	ctgcccgcct	cggcctccca	cagtgtctggg
1981	attacaggcg	tgagccaccg	cgccccgtct	acctttcgaa	ttttttactt	ccccatcct
2041	atccataatt	ttatacaaat	ggtttggtgt	tctctaattt	ttcatagaaa	actcttaggg
2101	tttatggatg	tattttctca	aacacttctg	gaatactact	tttaaaaaaca	tgcttttctt
2161	gtctctaaat	tatctctatt	ttctcttaga	gccagttttc	aaactttttg	ttttgtttta
2221	catcttggtc	catctttttc	atgctgtaag	gatttcttgt	gttttggttt	ttttaaaata
2281	tctggtaacc	cttggctgtc	ctttcatttt	ttcttttttt	tttttttttt	tttgagacag
2341	agtcttgctc	tgttgccccag	gctggagtg	agtggcacia	tctcggctca	ctgcaacctc
2401	tgctccggg	gttctagtga	ttttctctgc	tcagcctcct	gagttagctg	gattacaggc
2461	acctgccact	actaggccta	atttttgtgt	ttttagttaga	gatgttggtc	aggtcgttca
2521	cgaactcctg	acctcaagt	atccgccgc	cttggcctcc	caaagtgtcg	ggattacagg
2581	agttagccac	cacaccagc	caatttcaca	tttttaaatg	agggatcagg	tcacttaata
2641	taaaagctgt	gtcgggttcc	cctcctagg	gatagctctg	ctccccatta	gtgccctccc
2701	caagacaaaa	cagccaacag	cactctcatc	acatggctcag	ggcaacacct	ggctgcgctg
2761	tgtttgagg	gagctatcca	ctctccacc	agacatggta	gtttcagctc	ttttaacctt
2821	gctatgctat	taggtgcgtc	ttgatatttc	attgtgtttt	taaattttcc	tttttttttt
2881	ttcttgtaag	acaaggcctc	gctctgtcac	ccaggctgga	gtgcagtggt	gtgatcatag
2941	ctcacttcct	gggtcctaac	aatcctccca	cctcaggttc	ccgagtggct	gggactacag
3001	gcatccacca	ccatgccccag	cctcattgtt	tttaattgcat	ttttccaatg	tctagttagg
3061	ttgagtgcc	tttcatatgt	ttattgtcta	ttcatatttt	gtcaagagcc	tgataagagt
3121	ctcctgcccc	ttatctgact	ggattgtctt	tattatcctt	attatgattt	gtaggcgttc
3181	ttcacatact	tcagatataa	atcctttttt	ataaatattt	ccccagtggt	attagcgtgc
3241	ctttaatgca	caaacttttg	aaatgtgcta	ctatttatca	gtttttttcc	ttgagttcat
3301	ggtttggttg	gttttttttg	tctctttttt	ttttgagaca	gagtctcgca	ctgtctccca
3361	ggctggagca	aagtgggtgc	atctcgactc	attgcaaaact	ccgcctcccg	ggttcacacc
3421	attctcctgc	ctcagcctcc	cgggtagctg	ggactacagg	cgcgcgccac	catgccctgc
3481	taattttttt	gtatttttat	tagagacggg	gtttcaccat	gttagccaag	atggtctcga
3541	tctcctgacc	tcgtgatcta	cctgcctccg	actcccaaag	tgctgggatt	acaggtgtga
3601	accacggssc	ccggcctttt	gtctcttttt	aaagaaaaaca	tctcttactc	caagaacatg
3661	gagccatcca	cctacgttat	cttcttttat	tattttgcgt	ttcttatttt	gacctaaaat
3721	ccatgtgaga	tgctctctgt	ctatagagtg	agtggagggtg	acaggcctcc	tttctttcca
3781	aattggcccc	acaccatttg	ttaggctgtt	ttcctcctct	tctctgaaat	gtcacttttg
3841	tgacacatta	gcgtccctat	acacataggt	ccaattcagg	acactggatt	ctcttcta
3901	gatgtatttg	tcctacagtg	cactgcacca	caccctcacc	cagctaagggt	ggccacggcc
3961	ttccctggcc	tgacactctt	ccttggtgtg	ggccaaggct	cccgggggcc	cttgcatggc
4021	caggggctca	tttccagctt	cacgcctctg	tggccttgca	tgggtgaggc	tcaccttgca
4081	tgccggaagg	cccaaagtcc	tgccgtgtga	ggaagatggc	gtgatcgtgg	tattcatcgt
4141	ggcccggtgc	tggcttctgc	tggaggtagg	cccagcggca	gacattctcc	aggctctgag
4201	aggggttccc	gatctcgatg	aggctcatgg	actgcagggg	gatggagaga	aatggaagag
4261	agaggttggc	gcctgggtgg	ccagtggtgg	gccagcagg	ctgtagtgtt	gacagacYcc
4321	atccactggg	aatctgttcc	agggtgtaca	aaggccctgc	ccggtcactc	tcaggacctg
4381	ctccactggg	gctccagtct	ttgacaggga	agaaaagtgt	ccccagtcac	actcacctgc
4441	gtgagtttcc	acacaaagtc	accgcagttt	tcagggtgaga	tgacagaagcc	aagctctacg
4501	tcactaaagg	cctcgcatag	tcccacggct	tacccatttt	ccagaagtc	ctgtgtttgt
4561	tctatgaggc	acctgttccg	tgcaaggcct	ccacagaagg	caaatgtcat	gtggacctca
4621	ccccaaagag	ctcacagttt	tgggagggtc	cgtggcaggg	gaggcgga	ttccagtggt

4681	gtgtgcaagg	ccagaggaag	cagatcacct	tatacgttcc	ccccgtgtga	aggtacttgc
4741	agaaaccaac	acaatggagg	aaaagaatac	gcaactgaca	tatttgggtgg	gtgcaaaagt
4801	aatcgagta	ttgccatgaa	aagttaattg	aaagactttg	gcatcaacct	aatagaaaaa
4861	cagccatata	aagaacaagg	gcatacacag	atgattccca	caggaactat	gcatgggaag
4921	cccccgga	ccatccatt	ccattttgtg	tcatgacgca	cacctgacag	ctaggggtgg
4981	ctcgggaccg	ctcgtcttgc	acactctctc	tgcacatga	catcagcagg	cactcccat
5041	tagacttccg	gactaagaag	cctgtcttgg	acaggaaaact	gagactcaga	gaagttaaga
5101	agcttgccca	ggtcatgaaa	ctaggaagta	ggagaattga	gatcggatcc	caagaaggct
5161	gaaacaaaa	tctagattca	actcactgcc	ctaaacttcc	tccttaata	agacatacac
5221	aatatcattt	caaaagcaag	gcttctgtga	gcggcaccct	tccttttttc	ctccacaatc
5281	ctcacagcag	gcattttggg	ttctgttctc	accatagact	ccaaacacccg	aagtagttcc
5341	ctccaagaga	accgagagcc	cctctgagtc	ctgggcatgt	tcctttcccc	tttaaggatt
5401	tatctggcca	taaaatagag	atgtgcatc	cttggaagct	acgtatgcag	gtgttttttc
5461	tcttgcataa	atgtatccac	aaagtagagc	cacgtatttc	tccttgtgtc	tataaataac
5521	aaaaagttgg	ccgtgcgagg	tggtctcac	ctgtaatccc	agcacttttg	gaggttgagg
5581	caggtggatc	acYtgaggtc	aggagtttga	gaccagcctg	gccaacatRg	tgaaacccYR
5641	tctctactaa	aaatgcaaaa	cttagtctgg	cttggtggcg	taatcccagc	aacctgggag
5701	gctgaggaag	gagaatcgct	tgaacccggg	aggtggagg	tgcaatgagc	tgagatcacg
5761	ccactgcact	ccagcctagg	cgacagagtg	agactccgtc	tcaaaacaaa	acaaaacaaa
5821	caacaaaggg	tattgttagc	aagccagtta	gttccagaag	aagggagcac	aattcaaggc
5881	taggttaagac	aagtctgtcc	caggggtccc	agtcaatgac	tgacaagatg	gcagaccaat
5941	tcccaagtc	caagcacaga	gttaattgta	aagactttat	tcctgtctcc	tgaggagataa
6001	gataactcat	tctaaacagg	tttccatgag	tccgatgaa	ataagtata	aatcagagt
6061	tgatgtctcg	tgactaagac	ttgtgggaca	ttggcataga	agtggacaga	ccagaggaag
6121	ctggaggagg	gcaaggggag	gtgcacttcc	tggtcccatc	atgaccagga	agagcatgtg
6181	actagctctg	gccaatgagc	tgtagagcaa	agtggggcat	gtctcttttc	ggttggagca
6241	tctaatttcc	cattgagacc	cttcagagag	ctctccttcc	tggcacgaga	tctggcaaca
6301	taagagacgt	ggctgtctcg	tcacctagga	cctgagtgac	gatgagaccc	gaatccctcg
6361	ccaacccaca	ttgaacatac	agcgtgagac	ggaaatcttg	gtcgttttaa	gccactggga
6421	tgcaaagtgt	gtttgtttct	gtagcataac	ctcacctact	ataattgata	caggaggaag
6481	ggagtgggta	tctgcatcac	atccaagcct	gacccctgca	gtcactctcg	agtccttgcc
6541	aggtagtatc	cgctctggg	tagccttgag	aggtcacagt	tctgaaggag	gaatcgggcg
6601	aaagagtgtc	ccttctcact	gtctccagct	agggggtgag	cttgggatgg	caacaagaag
6661	aagggaaact	cggggtggat	gtgattcggg	gctagagcgt	gggccttggt	gttgggtgta
6721	ctcgtgaac	ttaagattgg	cagcccagat	atggatgagc	ctgggggacc	ttatgtctaa
6781	tgaaataagc	cagggtcaga	aagacaaata	ctgcacgctc	tcactcatac	gtggaagcta
6841	aaacagttag	gttcacagaa	atagagagta	gttccctggg	agtggggggc	agcgggctgg
6901	ggattgctcaa	ggctggttaa	catatccaaa	agtgcggcaa	gatgggggaa	taagcgttag
6961	tgccctatag	tgccggaaag	cgcttggaat	ccgtaagaat	gtattgtgca	ttttcaataa
7021	gctagaagag	tggaacttga	atgttcccaa	cacaaagaaa	tgataatgtt	tgaggtgaca
7081	gatagtctaa	ttaccctgat	gtaatcatta	cacattgcat	acacgtgtct	aaatatcaca
7141	ccctacccca	gaaatatgta	caattattac	atgccaat	aaaataacag	taagagcaaa
7201	atcaaaaaac	aaacaagaat	gactttttta	catacatttt	ttaaaaattg	ccagtccaaa
7261	tccaaatcag	aggggtcagc	caaaggtcac	taataagcgc	ttttgctagg	gtgggtggcag
7321	ggccctctgt	gcagtgtgag	tgtctccatc	caggacaccc	agcacagctc	ggtagccttg
7381	ctgggctcca	gggaaagggg	ccttgcttcc	tccagggcac	gtctcccagc	atcagtgtca
7441	agggagtcc	gtccaagatg	aactcacctg	ccctcagggt	cctaagcccc	atgtagtggg
7501	acgggtccct	ccgatgccct	gggtgggtgc	gggtgacttct	gcctgcccct	gtttcccttt
7561	cgtgaaactt	ccctcccttt	aatctgagtc	agaccccagg	cttcaggggc	aggcaggcct
7621	ggccaatccc	agcaccacgc	cttccctggc	accatgaagg	atcagggtga	acatacaacc
7681	aaagcgggcc	aaggagtagg	caaggtgtga	gggtgggggg	ctgcagcggc	tggaagcccc
7741	ctgcactaca	ggagaccccc	caccaaggat	ttaaaggagga	tgagaccact	gtcgagctaa
7801	gactggggct	gcagtctcat	ctgaaggctg	taccggaggc	gggtcagctc	ccaggctcat
7861	tcaggtggtc	acggcgggcc	ccagctgagc	tggtgcacac	cataccgtat	gggccccctc
7921	tggggcccgt	acccatgtca	gctttacccc	acacggctcg	ccctgctcga	gggacccaag
7981	agacaacaag	agaaggcacc	tcccctacag	aagccacagt	ctgcttttga	tagaatctca
8041	gtcccgccca	ccgcagggca	tggaacttaa	gagactgggt	cactggcggc	catcacgaag
8101	gctgctctgc	actgacgggt	gtcatttatg	cctgaagaag	ttcctctcct	tgtaaaagcc
8161	aattctgagt	tgtgtctctg	tcacttgtag	cgaagagaa	ctgagccacg	tgcttctctc
8221	tttgggggac	cacagggacc	ccagtcagga	cactcgctgg	cccatgaaag	atggcgctct
8281	atctgggggc	agctgatatg	ttctgggggc	agtttatatg	cttctcagtg	ggagaagggc
8341	cctggccaag	tctttgagga	ggagtcaaga	gtccagccag	aaagacatgc	cccaatatca
8401	ggaccagagg	caggaatgag	gctagtgtgt	aggacagagc	tgcgggagga	gccagctgga
8461	ggcaggctgg	gggtgtgcag	ggaggtccgc	tatcctagtg	tggtggcttg	tggggagtag

```

8521   cgggaggatg   ctgaaacccc   aagcgcctac   aggagcaagg   caggtgacaa   gtcaatgagg
8581   caggctgggt   gcacgaaaag   taagtccatc   ttgtgttttc   ccagttttga   aggagaaaag
8641   gtacagagga   atatttccct   attagaatag   agtgggggtc   ggtgataagg   agtggcgagg
8701   actgtggcaa   acccagggtc   ccaggcacag   ctgaaggagg   caacatccgc   ccagtgtggt
8761   caggaacaga   ggcccagagc   tgctggatca   gatttctcag   gtcaatcgtg   aaatctggat
8821   tctacagaca   cctacctaatt   agtaaacact   ggctcaaatt   gaaatcagtt   atgcagacaa
8881   aacaacgcac   cctccccgtc   accactttat   tccagaaccc   tggaaataaga   gccctgcagt
8941   ggKcagcccc   ttctccacgt   cctgcagact   ggatccactc   tggccgatca   cctctctcct
9001   ccagatgct   ggcttctgtg   gttcagcctg   ggttcccagc   ccgtcttagc   agccaaagtg
9061   atcttgctgc   tgcggaaga   caatttattg   aaatattgat   ttgtgttttt   gtctctgcct
9121   cgctttgttc   tcggcccagt   ccaagaaggc   agcctgccac   tttcgttctc   caaccccaag
9181   gaggttgtgc   tgacgggtcc   agagctaaag   ggggtctatga   ccccttctctg   gctttggggg
9241   tctgaggaga   ctccctgagg   agggggcaga   ggggagggtc   ggggtgcatg   ggcacaggga
9301   ggggaagctg   cactggggcc   aggggagccc   tgggaggcag   gacttcagga   gacatgagag
9361   tgtgctagac   cctgggagcc   tgggtcccag   cagggggtac   cccaagaagt   agaggagagg
9421   gttacctcag   tgtccgggca   ccccaggact   caacctgcca   ctgggagctt   aactgacaag
9481   ccacagagct   gcttggtgct   cctctctctc   tctgccaacg   tcccaacatc   tccaagggtc
9541   cagcctatct   ctgacgggca   cagaggagcc   caggggcaga   gatcttgtcc   ttgctgggtc
9601   tggagcatct   aagggtatg   ggttaccatc   gaggccagga   cattgacctt   gtgcccagag
9661   gccaggggcaa   ggtctgtgtg   cattcttaaa   acagcccttt   gggtcagtga   ggaaggagtg
9721   gggatagaag   gggacagtgt   ggcactgaga   tgcattgcagg   tgtggctggg   gccacgtcag
9781   gagggtcagg   acctggtggc   aggagctgag   aggcattcagg   gctgcagggg   gggagcaatg
9841   aggacagagac   ggcaccagtg   tctggctttg   gggattggta   gcaaggcaga   agaggcggcg
9901   cgggaagggg   acggcaagtg   ccagaggaag   agcgtggggg   ccgcagtgcg   ggaacatggg
9961   tcacggcaat   gtctttctcg   gccagcttgc   tcgttgggaa   gagctgggct   gttttgcttt
10021  gggagctctc   ccacttctgg   ggggtgcagcc   ctgttgggtg   atcagtccca   agccgtagct
10081  gggacccatg   ttaggcccat   ggagcacagg   ttcgaaaagg   ctgaaaactg   gctgaagtgg
10141  gtttgccctg   tggcagcacc   caaaacaacc   atccgcaaac   tcaggctcat   gcaccccgag
10201  gccccctttc   tcccaccac   tcccaccac   ccatgagagc   tacaggcagc   ctgggttct
10261  gcatgctctc   aaggactcct   cacagaagaa   aactgatgtc   caggcagaga   catcggggca
10321  gctggacagg   tctgcgcccc   aggagggaag   tctgggctgg   ccgtggacag   ggcaaggccc
10381  tcgtactctg   aggcagtcgg   aacccaaggg   aagagccaag   ctcagggtctc   tccgtcgaga
10441  cctccagcca   aggatactgt   cctgacatgc   agtgaactcc   cctgtgggtc   acttgggtc
10501  cagcgggacc   acctgctggg   atgcgacctc   ctttgcctct   ctgcccactg   gctgtcagtc
10561  aggaggccag   atgggtccct   gctgctggg   gcaccaccagc   cccaccatc   ctgtgccctt
10621  gctatgtatc   aagcctggga   actgtgcttc   ccctgagcga   gaagaagaca   catttctcat
10681  atatggatct   cctccagaag   tgaggggagg   tgcactgctt   gtttccactg   aggaagtctt
10741  caccagctcg   ccctggctcat   ctacgtccct   cccagtcctg   aatgcacctg   acttccagct
10801  gcctgtcctg   ctctgcacgg   ctctcttagg   aaagtctgtc   tgtgtgactc   cacctgttcc
10861  tagctgtgct   gaggggggga   ccaggagtga   gccgtagcac   cctgggctct   gcaagggggg
10921  tggctgcagg   accagccagg   gactttacag   ggcagccagt   cctcccagg   ggcagcccg
10981  ggcacccagc   ctccagagca   gggggcggtg   gatggtgctg   tatgaccggg   gatggacatt
11041  tcgaggccct   ggggctgcat   ggtagctgca   ggggtccacct   cctgagttag   gcagggcctc
11101  tgagggcaga   ctcttcacag   cggcagcagc   tcccagaact   cccaggtcca   gccatgggcc
11161  ccagaagagc   caggcgagg   aagcctgcag   tagggctgtg   gatgaagtcc   aggggcccac
11221  ccagcccagg   atgcaggagg   agtctttagt   gggctgacct   tagctcagag   cacgggaaca
11281  gatgagcttc   caagaggctg   atcggtacct   ctacaggggc   cccaggacag   tcgtaccccc
11341  tcatggaggg   ctctgccacc   cctgtttgga   ggtgacatcg   accactgcag   acatggggtc
11401  ccaactgggc   tccactggcc   gagtgtctct   gggccctgga   ccaggctggc   tcttccagag
11461  tgcagggtcc   tcttggcca   tcatgagtgc   ccacctatgt   ccacatggag   acctccttg
11521  ctatgctggt   tcacagagct   gccagatgc   aggaggcaga   cactcgtcca   ggccccctgg
11581  gccctcccat   agcagcaccc   ggagcctggt   cccaagaagg   atggtgtttc   ctccgggttc
11641  agcctgtttg   gctcccctgg   cccgggctgg   aaagggtctg   aggacatgcc   tgaagagccc
11701  agggatccca   ccattgaggt   gagagggcag   atctgagggg   cttaggagg   caagagcttg
11761  atggccctca   ccttgggtca   gaaggtcact   ttgtgaaatg   ctccccagg   ctgtcaggag
11821  gcttggccag   ggcaagggtg   aggaaacact   gcggggccct   gaacctcagg   cccagcctgc
11881  aggggtctgg   gtaacccac   gggcctcagg   gaacagccag   gagcttaggg   ggccctggtg
11941  ccttgggtgg   acgagtggc   tgctgccacc   tccaagtccc   agccccatcc   tggccgcggg
12001  accctggctc   tgagctggcc   gcccgttccc   tctctctgct   tgcgatcggt   cacctgccag
12061  atggggcacc   tactgggat   tgctctgttg   ctgtcgctgg   cctctgagtc   actaaagcct
12121  aagctccttg   agggccgggt   ggttgcctgg   cccgagctct   tgccccagcc   ctggccccag
12181  cctgcccctc   gtacccccac   tcttcccctc   acggagcaat   cccggcttct   gctctccatg
12241  ccaggatgag   gcagcgaaac   cggccagagg   accccggctg   ctggggcgct   agtgatgaag
12301  gaatgaacac   gagcacctc   aggagccagg   ctgggcaggg   agggaggcgc   cggctgcggg

```

12361	gacgttctcg	ggaggcctgg	ggaagtgagc	ccaaagggcc	gccccggcag	atgctgaccc
12421	cagcaggaag	tcaccgtccg	caaggagcgc	cccgcctaggc	tcaggcttgg	cacacgacac
12481	ctcgcttcca	cttcattttc	ctaccttccg	gatgacgaac	tgcgtctcag	acaactgaaa
12541	ggactcgcc	gagtcacaca	gccggaaaga	ggtcgggggtg	ctgccccccg	cgcgggggcg
12601	gcagaggaag	ggccgggccc	gagcggccac	cagggtggcgc	tgctgccgcg	agaacagcgg
12661	cccggggccg	ggacgcaccc	ggcgcccaga	ctccgcggcc	gcacccgggc	tcccggggccc
12721	ccagcgggcc	caagacagcc	tctgggcttc	accctaaata	catcaattat	gacctgagca
12781	aaaagtctcg	ttccaatcgt	tgcctccga	catttccatt	gtttgctctg	gaaggagctg
12841	aattatctat	gctcttccct	ccttcagtca	aggagacgtg	ggaggaaacc	gggatgggga
12901	tcagcagcct	ggaggccgcc	tccgctggaa	ccgcgcccgc	cccgcagga	ggctctgcct
12961	gctgggcaag	gcgcactcag	ccagccccgc	aggccgccac	gcgaaccgag	ctaccaaggt
13021	ccactgcagg	tggcaggtga	ggtcacctgg	gcagctctag	gccccggcga	aatatctcat
13081	ctgctctgtt	cctgggcaag	ttaccttact	tgtccccacc	tgcaaaatgg	gcacaactgc
13141	agctccgagc	ttggagctgg	ggggcttccg	cccatgacag	ttcaggcagg	gtgcagtcag
13201	cgcccacaaa	tgtcccttgt	ccttgtctct	gccgtgctgt	tgatactgat	ggagtataac
13261	accagtccct	ggggactggc	ctacccgctg	gggctggctc	atgcccccaa	ggggatgcgg
13321	cttcagcaaa	gcttctcaac	cgtecccatc	ccagaggggc	cctcctgcct	cagaccagcc
13381	tcctaaccag	gtgcaggata	ggacagcagg	gaggttgggt	tctgggtgagc	agaccacatt
13441	tctcaacctc	agcactccag	atatttgggg	ccagataatt	cctgttgggt	gagggcgga
13501	ccctgttcat	ggcaggatgt	tgagcacacc	cctgggcact	ccttgccaga	accagtgga
13561	accctcccc	tagtcgtgac	aacccaaaagt	ctctccagac	actgccaat	gtcccccg
13621	tggggggagg	ggtggcaaa	tcactcttgg	ttgggagcaa	ctgacctca	ggggaccagg
13681	acaaggagcc	tgaagctggt	gctcctgacc	agccgggac	acagagtcca	gccctctctc
13741	ccagagcaga	gtggctccac	cagcgtgtgg	ccccgggag	gccgctcgcc	ccctggagcc
13801	agggtcctca	ctgtcaaatg	gggatgatca	cagcgcctac	tccacaggtt	tgtctggggg
13861	gtgagggcat	aaagtacat	aaagggttc	agacggtgcc	tgtcccagaa	gaagctctac
13921	aaacagtcat	ttgcacgaac	gcctttatga	ttgtcgtcat	tctgttgtg	tgttgggggt
13981	tcacagacct	cctcttatgg	gggaaaacta	ggccccatga	tgttaaattg	ctcaccaaag
14041	gtcccacgag	tatggacaaa	gccagaatga	atggcaggga	caagggtcca	ccatccctgc
14101	tctgtgagca	catgttctac	cgtgcccccc	agctgagtgc	aaagtgggct	cccccaaggc
14161	tggaggtctc	ttgttccagg	cagggtttac	agatcaaatc	tcatcacgac	agagaaccac
14221	gagagctggg	gcaggaggcg	tgcacctcat	ttctctgcca	gaactcccat	atcgttccag
14281	cattttgttt	ccagccatct	tccccaaagc	ctgggaagct	attcccagag	agagccaagc
14341	tgccagaaca	ggattctcaa	cttgagcctt	cccaggcatg	agccagtgca	ggcctgtggg
14401	gaggcagaag	aggggcctcc	tgttagaatc	tgaatgttgt	gtccccca	tttgaaaggt
14461	tgaaagctga	gaaccaatgt	tatggtatta	ggaagtgtgg	tctttgagag	gtcatagggt
14521	catgaggggt	gagtcctcat	aaatgagatt	agcaccctta	caaaagaagc	ctgagggagc
14581	ctgttcaccc	ttccatcatg	cgaggggaca	gctagaaggt	gacctctagg	aggaatgggt
14641	ccttacgaca	catcagttct	gctggcgctc	tgtcttggga	cttcccagcc	cccagaacag
14701	tgagtaaac	aattccattg	gttataaatt	actcagctca	aggtatttct	gttatgagc
14761	cccaaacaga	ctaaggcacc	tcccaagggc	agtgtccagc	cttagtgctt	cctagttctc
14821	tggaccacgg	taggtctggg	tgaggagggt	gatggaggat	gatgtgggtg	ggccatcatg
14881	ccttgacccc	ctatcaggtt	gggggttggg	accagatgc	tcaggaggac	aggggctgtg
14941	ccctgggggtg	aggcatgggg	tgtgtttccc	acctcatccc	ctgtcccaga	tccccagggc
15001	ccttgatata	tctccccag	accttcaatt	cctggccttc	ctgcaaccatg	gctgggccac
15061	tctaaatcct	ctaccctata	ggtgggtggg	tgggcacacc	aagccccac	ccccagcacc
15121	tctcttctcg	tgtttccac	taagagactc	acatgcagaa	tttctcagac	ctttcacctc
15181	caatgctgaa	aaattccagc	aagcattcca	cactttgtct	tgacctgggt	ttttataaca
15241	agacacttta	tgtccttctc	tgatgctccc	caagtctctg	atttccattt	gctttgat
15301	tcatcatcct	aggaaatatt	taaaatgaga	acagttgagt	agctttggta	ccttctcaaa
15361	aaatcaattg	aaaagccagt	tgtatgcaa	cctcaagtgt	tgcgatgct	ctgctccctg
15421	gctgggtgga	tattgtataa	aataagatca	tctcttttct	tctgctttt	tttaagaaat
15481	ggagaaaaat	gttgttttag	gaaaaacttg	agataactca	gaggcttctc	ataagctaga
15541	agacaaacat	gtcactcaaa	caagtgtctc	caattttatc	ccactggag	cctggtgagg
15601	cctgggggac	aggcaatggg	ggagcggcag	gtggggagct	ggtctatttc	tgtcacaatt
15661	tgggatgaat	aacaggaacc	gagtatctgg	gccactttct	gccaagtct	gggtcatggg
15721	ctcacatccc	ccagagtctg	tccccacctg	gggggggaca	gctcagcccc	ctcctgactg
15781	ccatgccttc	ggtttttgcca	tgttaaaatt	ctctagaga	cgggggcttc	attccttgac
15841	gggaagcctg	tggttccttg	aagagcagct	catcagaatt	catccccacc	atgtgccaaa
15901	ctttgtcccc	agagcccat	cgagaatccc	tggctgtgct	tctctgtgag	caaaaggggt
15961	ctgttccctt	ggacacagga	gctagacctt	ttgccccag	atcttggtg	ggaagtcccc
16021	atcctggctg	ctccatccat	gggtccctat	tgcctgccac	cctgctgtat	tcttgcggg
16081	tgtcaccac	ctgtctgctg	tctgtctgct	tgtactgcc	tggacacagg	gctcctctcc
16141	cctcctcggt	catgcacctc	cctgggatta	aacagcagct	gctcggaag	ccagaggctc

16201	tgggctggcc	ggatgctttg	ggggcacaa	gatacatctc	ctcccacgtt	catcagacac
16261	ctcagagacc	tgaaaacagg	gcctcccca	gtgctatgca	aagtgaagtc	tgtgcctggg
16321	ggacacaaac	tgttcagtc	cacagtga	taagtacaga	aactagaaac	ttttatagca
16381	attggacggg	ttgagtttat	gtccacagaa	tttaataata	aaaatccaga	cctagagttt
16441	ttggatacct	cggtttttgt	cacatttgct	tttccagta	attcatttat	aacgtatggg
16501	acggaagtat	tggctctgtga	tgggtttgaa	ataggaaaaa	gaaaacagcc	aggctctttg
16561	cattgaggtg	cactgtgtac	tgctgtccc	aggatggcct	ctcccctggc	cacctgggga
16621	tggcagatgg	ctcctccgg	cccacagacg	ctgctcccaa	gccctgtggc	ccacagcaca
16681	ttcgctaagt	cagccagggc	ccatgacag	aacctatttt	cctagccgct	gtgccctcta
16741	ccccagcct	tcagcccagc	ccctagcagt	gccacctcct	ctgctctact	atttcctgca
16801	cttttgagct	aggccaacc	taaagcaacg	aggactcata	gcctcggctc	acttcccctc
16861	tcacagccca	gaacccccca	gagaagggca	cagcagcggg	caggcccggc	tcttctggca
16921	gtctacactc	agctgatcaa	ctcaggagat	ccccgggatg	gagcggaccc	catgggcact
16981	ggcgctccct	acccccatca	ctgcctcctg	tgcttgatct	gggaggcttc	taggaggaag
17041	gcggggtagg	gccaggcaga	aaaagggtgg	cagagtatgg	ggtgagcttt	gctgagcccc
17101	actctgggca	ggcatgagtt	aggagcagaa	aaaaaggccc	agaatgctgg	aatggtccag
17161	aacattccag	gaggccaatc	atccctgaat	tcccggggtc	tgacagaggg	accggtgtgt
17221	tgttgtcctc	caagctaggc	tgaggggcct	gggaacaggc	actggggtgt	cctgggtccc
17281	acacccctca	caacccagca	ctgcctcctc	cctgcctcct	cagctgcggg	tccactgagc
17341	cctcggggcc	gggggtaagc	tgggtccatc	tggagttagc	tctggaagcg	tgcgccagc
17401	ttcgggttct	gtcttccaat	ccctcaccac	catcagttct	caagcccctt	cccagcagga
17461	ggcggaagca	aggccagcct	cagttcacct	atgtcctctc	tggtctcttt	gtacctgggt
17521	cgagaagtt	gcccgtgccc	caccacaggc	atgagctgcc	ctgggttgct	ggctatggat
17581	ctcagcacc	ccagggcctg	cacatgtccc	catgttgctc	ctgggaagag	caggctgcac
17641	ctgacaaccg	gtcccctgcc	cactgcccga	ttcccactc	tgccaaaagt	ggaagtgggg
17701	ccccagcatg	gccccagtg	agcagagctt	ctgcagggtg	atgaggggtg	ggggtcccag
17761	caaagcagaa	cttctgcagg	caaggcgcat	gaggggagg	ggccaaggag	gccctggcca
17821	ggagggtggt	ctctgctaac	ccaggggcgt	gggcttctca	tgacggggag	accagctacc
17881	ttgggcagcc	tgcgctcctc	ctccaaagt	tcccatcttc	ttcagggtga	cccgctgcag
17941	gtggaccccc	tgccccagcc	acaggaaRcg	tgagctagcg	cccatagggt	ggaacacaga
18001	aacctcatta	ctcctcccca	gcccactgca	gagatgacgt	gccaaaggaa	agcatgaatt
18061	ctctttgatg	tcttcattga	aatcagaaat	aaaggactga	gcgtgacaga	gccctgtggc
18121	atgccaccgg	agacctttct	ccagatgtct	gaattcagtc	aaacggcatc	atgtggatcc
18181	taccattttt	aaatgatttt	ttaaaaaatc	atttaaaaaa	tttaaatctc	acaaatttaa
18241	aatgcagtg	tttggttata	aaagaaacat	gctgggtccag	gctcagctgc	tcacacctgt
18301	aatcccagtg	ctatgggagg	cagaggcagg	agaatcattt	gagcctggga	gttcaagacc
18361	agcctgggca	acagagtgg	accctgtctc	tacaaaaaaa	aaaagaaaaa	aaaattagtc
18421	gagtgtggcg	gtgcgtgccc	gtagtctcag	ctacttgagg	ggccgagatg	agaggatccc
18481	ttgagcccaa	gagatggagg	ctgcagtga	ttatgatcac	accactccac	tctagtctgg
18541	gcaacagagc	tagaccccat	ctcttaataa	taataataat	aataataata	ataataataa
18601	aaaccatgct	catcatattg	aatttaaaaa	attcagtaaa	gtacaaagaa	gaaaatgaaa
18661	attaatttag	ctactgccc	acagaaaatc	ctgttcacat	ttgatgtgtt	ccttccagtc
18721	ttttgggtaa	atatatttgt	acacaaaatt	gagatcatac	ggaaaactga	gatctgctct
18781	ttcaaaacac	aaacagcttt	agactgtaac	catttcccaa	tgccattaaa	catggtttcg
18841	aaatgcaatt	ttgtggcagc	ctaatactct	accacatagg	tttatcatta	cttaactaga
18901	ccactattat	tggacaatat	aggttgtttg	ctgttggttc	catgagagat	aatgccatga
18961	cccttgtctt	gttcttaaaa	cttggccccc	agctgggagc	actgctttgg	aacagagttg
19021	tagaaaatga	tgtctggggt	cacagaaatg	aatgtctggg	ggcaaacatc	aagaacattt
19081	taaaaggcaa	atgcggtggc	tgcgccccca	taaccccagc	actttgggag	gccgaggcgg
19141	atcacctgag	gtcaggagtt	caagaccagc	ctggccaacc	tggtgaaacc	ccatctctac
19201	taaaaataca	aaaaattagc	tgggtgtggt	ggcagggtgc	tgtaatoccc	gctacttggg
19261	aggctgaggg	aggagaatca	cttgaaccca	gggtggcagag	gttgacagtga	gctgaggtcg
19321	cgccactgca	ctccagcctg	ggcaacagag	caagactcca	tctcaaaaaa	aaaaaaaaaa
19381	agaaagaaaa	aaagaacatt	tttaagggca	gtgataaata	ttgctacatt	tccttagaat
19441	atctgtactc	atactcaaag	tgtgcctgct	tcattcaccc	agccataaat	ccctcttcac
19501	cttaacaggg	ttgatgcttt	ataaatgggt	tattatattt	ttctttaaca	gtagagttag
19561	atatttttct	atatgggctt	tctcttagcc	acttggattt	ttttctttgt	ttccttccat
19621	tgacaatttt	gtttccttat	gattttaggg	aggttttata	tatttaagata	tatttatcac
19681	attgcaaatg	tttttctctg	ttgtcatttt	ccttttcatt	gtgttttttg	acagaagttt
19741	ggtattatac	ttatccaaat	atatccatat	tcccctttac	aattcccttc	attggattta
19801	tgcttagaaa	gctcttgaca	accaatattt	tcttcttgat	ttgtgtgatt	ttaatagttg
19861	aatgttgtat	tggcttatat	ttgtcaYatg	ataatttaat	ttcctattta	ttaaatttat
19921	gttatatat	tatttataac	acaataagca	ttctgtgttt	atcacggtag	gttatttgta
19981	agaggggtat	tttcaactat	tttattcaac	tgaactgtct	atgaaagttt	atgcacatat

20041 ggcactatatt taattatctt cactttacaa gatgtttgga ggaagtctct tacttgcttg  
 20101 ggtcccacag ctgtctcggc cattcttgac cgttcacgtt cctttggagg aagtctcttc  
 20161 cttgcttggg ttccgcagct gtctcagcca ttcttgacca ttcatgttcc ttggaggaa  
 20221 gtctcttccct tgcctggggt cgcagctgtg ctcagccatt cttgaccatt catgttccct  
 20281 tggagggaagt ctcttccctg cttgggctcc gcagctgtct cagccattct tgaccgttca  
 20341 tgttcccttg gaggaagtct ctctctgtct tgggttccgc agctgtctca gccattcttg  
 20401 accgttccatg ttcccttggga ggaagtctct tccctgcttg ggttccgcag ctgtctcagc  
 20461 cattcttgac tgttcatgtt cctttggagg aagtctcttc cttgcttggg ttccgcagct  
 20521 gtctcagcca ttcttgaccg tccatgttgc tttggaggaa gtctcttccct tgcctgggtt  
 20581 ccgcagctgt ctccagccatg cttgaccgtt catgttccct tggaggatgt ctcttccctg  
 20641 cttgggttcc gcagctgtct cagccattct tgaccgttcc tgttccctaga tgggtgcttag  
 20701 aatcacttca taacatttcc aactatgaat ctgggaaagg ccaacctctc gattataccc  
 20761 agcttcccat gaggacagca ttgtgtctct catttgtgat cactttcttc ctatcattaa  
 20821 tcaggtacaa ttctctcca atgggtccat tgtcttgcga ggactattcc aggtattcta  
 20881 catttttgggt tactaccata atggaaattt tctctccctg ttttcccccac taccaaggct  
 20941 gttgtttttg gttcacacag aagctactga ttttttaaac aatgggtctga tatttggcca  
 21001 catcaccaga ctctctcct aactctacaa gttRtaccgt ggattctttt aggttttccc  
 21061 gacacatgat taacagcaaa caatRagaat ttgtctctgc tcttttccaa tactagtttc  
 21121 atgttcacata caagtatcta ggattttcag aaccacagta actagtttat ggggtttttt  
 21181 tgtttgtttg tttccaagac tatctccagg gcagtgctcc ccagagcagg ttcagcaggc  
 21241 tgcggatctg ctgagcaaga agcttccgtg gagaaagttt tcagagacca cctgtactgc  
 21301 tccgactctc ggaggtacac catgcacagc tccacaggaa agtctccggg aaatttccac  
 21361 cccaggaatg cccattttcc tttgtttgac cagcattttc cagacttctt tgaccttggg  
 21421 accctctat agtctctgt aacatctact catatcctaa gaattcgctg tgcagaacac  
 21481 ttttagggaa atgtgcttct ttttccgct ttaaatacaa tcttgggtgt cctgtgccc  
 21541 tctgttacta gttttctggg gatctggaaa aggaggagtg cctgtgagg ggctaYgggg  
 21601 ccgtcacgcc cttgccctgc cagaggcttc aacactggag ccttgaggcc tctctcccag  
 21661 gctcccccta ggcagcaagc tccctatgcc tgcagttgct tctgctaccc ccagctccag  
 21721 accccgcagg gctccagacc ccatctcaca tcttcttggc ctctccgct gggatggcac  
 21781 caggtacctc taactgcaca tctcagctga cctcttcagc tccctcctg tctccccaa  
 21841 tctcaggggt tgataccatc atctaccag ctgagagtct ctccgcttc acccacctag  
 21901 tgccacatcc tctcgaaagt gctcctctg tctcctgtc tgtatccctg gacccaggcc  
 21961 tggteccagt ccagctgcca gcatccccca tttgatggca gtgactgtct cctaagccac  
 22021 tgctcccacc gtcctctct gtcctctctg gccagtgatc ttcttagaaa tgcacatggg  
 22081 atcgtgtcac tgcctaaagc tgccctcatg ggcaagccaa acccttggc gggagacgca  
 22141 ggtctgctct gctccttgcg atcggctcct tccgacgtct ttggcatttg gtctgctgtt  
 22201 cctgtccctg tgcctgtccc tctgggttca gctgccctcc ctccccctcc agcctatgaa  
 22261 ctccagcat ctctatgctg gtgacttcc cctgaccttg taaggaaaga ctgacctgac tctgtctcct  
 22321 cctgctctcc cggtgcttt cctgaccttg gtacaattgt tctctgtct agatctgct tctggaagcg  
 22381 ccatccagcc ctgagccctg gtacaaattgt tctctgtct cacagcactt cccacacaga  
 22441 gtccctgag gacagagact atgccagtga atattcccag cactccttct cctacctgag  
 22501 agcccttcaa acagtatttg ttagaggatg agttttgtca ttgggaaact gatgtgatga  
 22561 actgtatgaa tagatttctc gatattaaac aatctccaca ttcttaggat aagccctaga  
 22621 tgcatacggg agaagtctct cttatataac tattcaatta aactcttaat gccttgggat  
 22681 ttttacttct attttttagaa gtgagacagg tctgtacttt gtctgacgg gctacctttg  
 22741 gcatcttttg gtataagggt cccattgact acgaaagaca cattgggtag atttcagaat  
 22801 acatgagttt acagggattt gtattagtct tgccctgataa atgcttgcaa atgagtatga  
 22861 atgactggc aaaacttttg aaaagggtgaa gaacttgagt tccctccatg gtatgggac  
 22921 ttatggcata ttctccattt tttggagtca gctttgaaat attcacatcc caggaaatcc  
 22981 tcaatttccc caggatctcc gtgtcccttg catagagatc ctcaaattct cacttccagg  
 23041 gtttgactct cttctgcaaa tgttctgact gtgatccat actccttcta cctacctgag  
 23101 gcagttaagg aaaaaaaaag aaaaagaaaa agaaaaggct tagggctggg ataatttagc  
 23161 atagagcctg ctaaggagta attaataaag gaggttttat tccatcaagt actcaagact  
 23221 gggaagagc agacacttt caggccaaat aagatccaaa gctttgccag cactggtaag  
 23281 caattaaacc tagtcaatta caaaattctt ctttttaaac aatgtaggag caacctgggg  
 23341 gaaaatgcaa gccgccacct ggcacccaga ccaatcaaKg ggataatctt ttctccctt  
 23401 tctagcccca ggaccaggag tttcccat tKggaaactt ccaaaaagagg atataactct  
 23461 gtccatctct catgcccagg gcttgagaat tcttaaaaca ccagaggatc ctggccaggga  
 23521 agggacagc tggaaagacc cggcttgggg aaggccggaa cagggaactc cactggaggc  
 23581 aggcagccag ccacatggga tagtccaaag atcccccccg cactcccccg cctgtgaatg  
 23641 gagagctgga accggattgc tgggctgggg cccaccccca ggctctgcat ctgtgtggtc  
 23701 caaagcccc tgccacaaac ctgggataga gggctcagcg agcctctctc tgccctccta  
 23761 gccagcattc actggcagag gatgcctgc acatccctcc tccactgatc atccctggac  
 23821 ctccgaagtg gccggctgca gagagcctcc caacagcttc ctagtgggca gatggccctg

23881	catgctttcc	tttctgccag	gagaagtatc	tttttgtagt	cacattacat	atactctttc
23941	tgatctttta	gtttgtgect	tctctctttt	tttctagatt	aaataggtca	gaggggtgct
24001	ttttccMaag	ctctcttctt	cattgattct	atgtgtttct	tccattttct	aattttattga
24061	tttctgcttt	ttctgaacta	tttctctatt	tttgctttcc	tttaagtttt	ggttttgttg
24121	ctgttgaaat	tcattcaaat	gtactgagtt	gaaggcaggg	ctaagtgtgt	tcgaacatat
24181	cacccaagga	agaagctttt	ggaaaacatc	tgtgaatttg	cctctgaata	cagctttaac
24241	cacattccaa	aagatttgac	agtagtgctc	ttgtttttat	tagtgttcca	aatagtcctg
24301	actgcagtc	taatatctc	tttctctgct	ctgttaccca	ggctggagcg	cagtgggtgtg
24361	atctcagctc	actgcaacct	ctgcctcccg	ggttcaagca	attcgccac	ctcagcctcc
24421	caaatagttg	ggactacagg	cacatgccac	cacaccggc	taatttttgt	attttttttg
24481	tagagacagg	gtttcaccac	gttagctagg	ctggctctca	actcctgacc	tcaagtgatc
24541	taccacact	ggcctcccaa	agtgtctagga	ttacaggtgt	gagcgaccat	gcctggcccc
24601	tagtagggtt	tttttttttt	taatgttttc	aaatgaattg	gattttttta	ggtttataaa
24661	ttctgctgcc	cccaccctta	cttttaggca	tttcatcagc	gaagatggac	aaaatcgttt
24721	tagctgtata	aaattcacaa	tattttcttt	atgacccaaa	acacttctgt	aatttttaag
24781	tgttctatga	ttcctcgaaa	agaaggttca	tctttctttt	tttctgtagg	gcaagggaaa
24841	cagtgatata	actactaatt	tagctttact	cattgtgtca	ttcaaaaatt	ctatgcctta
24901	ttagcatcta	atctgtgaaa	ggcttttaaa	agtgtattag	agactctcat	agttaggttt
24961	ttgccaattt	cccttttgtt	ccaaggattt	gttttgggat	gttttgtgtt	ctcttggggg
25021	acacgaaggc	ctgcaactgg	catccgccct	gtgggtgcct	tcgggtcactg	tggaaatgagg
25081	tgcttgggtg	gtgtagggaa	gtggccccaa	atctacttag	tattcctggg	atcccttctg
25141	ccaccggagt	ccttcgtcat	catttttaca	tcgtgtgect	tataagttgt	atgcagtttg
25201	attttatttt	tttagactga	gatttctttt	gaacagagtt	caacttttat	ataattatct
25261	gatgtatttg	atcctattat	gatttgattt	tctatgcaca	gccagaccac	aggctcgaat
25321	tctgctgtct	gggtaagaag	cctgactcct	cgggtgtctg	atggccctgg	ctgcttacct
25381	gggcatgtca	ctaagtccaa	tttaagcctc	gtgttcttat	ctgtgaaatg	ggcagagtca
25441	cagtgtgcac	ccctccatct	gccctgtaga	tttaagggcat	tcateccatac	aaagtgcctc
25501	ccacacgggc	ggtaacaggc	ctttgctgct	tacacggggc	gtaccgtggg	gctcctctct
25561	tcctctgtct	gcacagcatc	ccgtgttgcc	atcccaggct	actgtccctg	gcggctacac
25621	acccatgcac	accgccagga	aaggacgtcc	caggtttggc	tgcaagagga	agtgggacgg
25681	gcaagagagc	ccagctggga	cagcagccga	ggccagagga	gaggaggcca	cagatttaat
25741	acgggggagc	aagtcacaag	gagggctgtg	tttaaccacgt	gtttctgatg	ctttgacatc
25801	cagaccttgc	tgactctgga	gggactggcc	cgttaagtgg	taaccaatct	ttagagacag
25861	gaaacaactc	atccatgagc	acatttttca	gatacaaaacc	aaccaatcca	gcgcccattg
25921	ccccagccac	ctcctttatc	agacttcaca	ctctgagcca	ctatccccct	gccctcatta
25981	cccgagaaac	taactcagga	ccgctccact	ccccagagcc	cactggaatt	attcaaaactg
26041	gccaacctta	agccagctca	ccctgtccct	atcgggtcctt	cctgcagaaa	caacatggac
26101	gtcaggctgt	gtttacccaa	Ytccctctgc	ctcctgacca	ccctgggtgc	ctccctgtgt
26161	ggcccccggt	gcctggtagt	tctcttctct	ttgggatctg	ggagtaacag	actcttttca
26221	gtgaacaata	gtaatggcca	catttttaaa	caaggggtga	aagggagcag	ggWtctgtt
26281	tccttatagga	acctcagggg	agccagggag	agggcagggg	gttctggcaa	acgcacacac
26341	tctccaagga	ggcccatgcc	tcactccag	gccccctact	ccgagggggg	ggaggcaggc
26401	ccggctggcc	acagtgactc	cacctttcca	tagctcagga	ggatgatccg	caccagagacc
26461	acgttgatgt	gggcacccaa	ggactcgtca	tggtagattt	cattgacctg	aaagaaacag
26521	ggaggcatca	gcgggaacca	caggccctag	gactggctct	ggctctgcca	atgggatgac
26581	ccccacctgc	tccttcttct	tcccatgctt	tccaccaggg	cgtaaccatc	aactaggagc
26641	caccttgtga	ccctccctct	attgccccct	gccccctgct	tctccagggt	gccaccacca
26701	cctgctggga	tcactgtagc	cccttccctg	gcaacccccc	ccaccagccc	agggWtctct
26761	cacctggggc	ctgaacatgg	aacgtgggca	gggaaagcag	ccctacgctt	gctgtagcct
26821	ccagctgaaa	cacaaccttc	ccttcaacag	tgaacgtggg	tgatggaccc	ccgcggcaca
26881	ggcagggctg	ggaccggccc	tggcagaagt	cactccaagc	tatagctggg	gggtgtgggg
26941	attacctacc	tgcatcgcta	cttggaaatg	ggggcggttc	tagggccacc	accaggtata
27001	atgttttaac	gtgttaataa	aaagctcata	ttactgcata	cccaatgtgc	tctccctgtt
27061	ctgttttgat	ggctggagct	cctgcagtcg	ccgtctctgt	atttatcaca	atcatattct
27121	tacacgtgtt	tcattcttca	cattttaaac	cattattctg	agcaggtgcc	aatgggcttc
27181	agcagatata	aaagggaccc	gtggcacaaa	aacagggggg	tgaactcccc	ttctacagaa
27241	cagagccagc	aagaccctcc	agaaacatag	gtcaggttat	cccgggctcg	aggctgtcag
27301	gggcttccct	cgactgagc	ccccaccctc	cagcttccca	ccgcacacac	ccagggccct
27361	ccctagcgca	gcccccttcc	tgtaacctcc	tttgttgggt	aaccaaacc	Yactcctcct
27421	cgaggccagc	tcagcagtc	cttccccctg	gaaggccccc	cgcaggcctg	agccgcccc
27481	catcaagcac	aaggtgtgtt	atctctggag	ccctctgtgt	gccccgtgtg	gctgggcagc
27541	cacctgcac	ctccctgctt	gcatgccag	agccaggcta	agcagcacct	ggcaggcact
27601	cagggatacc	cgctgaaagg	ctgaagccat	tggcaccatc	aaagaagtga	gaaacagga
27661	ctggctgtgt	gcagaccgag	cccaggggct	gctgacagct	ccggacatgK	ccaagacata



27721	cttcccggag	cctgctgcca	cgtcaggcca	gcacacactc	cctgcactcc	aatctgtaga
27781	gcctgcatct	gagcaacagt	gattgctacc	tcaaaacagt	gggaaggcac	ctcaggaaac
27841	cYgagggttc	acaaggagaa	tgacatcatg	atggatgggt	ggacaggcat	acagggcggtg
27901	caagacaagg	atcccagagg	ccgctggcat	cggtgctggg	aacaaagatg	gcaccgatcg
27961	tggggcagag	ttcatgacat	gtggaccctg	ctgagggtac	catccagccc	aaacgcagct
28021	cctcttggca	actgctgccc	agcaaatcta	catctgat tg	taaattgggt	ccccacaact
28081	cccacctctc	tttcctatag	gcaagggaca	actcatcaga	cccacgggag	cgtaggctgggt
28141	caccgtggct	ccgggctgta	ggcgctctggg	gaggggtgggt	cagcaacccc	tgtaggcgct
28201	gagggttatc	ctctagaccc	ccagtgaacc	atctcgcccc	tgtaggctgac	aagatggagc
28261	tccgaggagg	gcacccaggt	ccccaggaag	ccttgccatt	tggtcatcgc	cctccgtttc
28321	ctctgtaaaa	tgaggattgc	agtcccccca	tcacaaagct	gttgtagggga	tgaaaccgYc
28381	atgcagggtgc	aatggaatgg	tgcttgtgca	aacaatgaaa	cggtttttca	ttgctggtaa
28441	accatagtgc	cctgcagctc	cctgtgaggt	cctgggggta	gggtatgcac	ccctaaacca
28501	cagaagtcct	tgcccagctt	tgagggtcag	actagattgt	ttagtccacg	ctggcacagt
28561	agccttaaaa	gccagttttt	cttaagagag	aaggggtttc	tgtaattgtc	cacagtattg
28621	ctttctctat	cacgacaaaag	cttggccagc	ttggctcatg	gagtctgacc	gaggggaaggt
28681	cccactgcag	ctgaagggag	cagaatagga	gagaggagaa	aggccaggga	tgaggggcca
28741	ggcttctcct	cctcctcaac	cccaacctca	caggtcatatc	agaacagtc	tgacattgga
28801	gaaggaagga	aatgcaaaaag	agcttgatcc	tggtgctgctc	ctgggacctg	agaggggaca
28861	ggttcctagc	agtaggagtg	gcttcacagg	cccaactggt	ggcccagcct	gctcaggcct
28921	gagaggggtg	cccagctccc	agctgtgtgt	agcagatgcc	aggggtccct	gcaccccaga
28981	gacacatggg	caagatagtt	ggggactggc	ttgcagggtt	acctggctct	agggggcacc
29041	accaggcttc	agggcaagcc	caggggggca	ctggcctaca	ggagaggagg	gcaggcctgc
29101	ttgcaaggct	actgcagagg	ggactcatcc	ctactgcccc	ccaccaggcc	tctgcctgca
29161	aatggggctt	ttgggaggta	attaggggta	gatgagttca	tggggggtggg	accctcatga
29221	tagggttagt	gtcctatagg	aacagacacc	agagagcctg	catcctctct	ctctgccgtg
29281	ggaggaccca	gcaagaaggc	ggctgtctgc	aaaccaggaa	gagaatcctc	actggggaat
29341	gaccagctg	gcaacatgat	cttgaatttc	ccagtcttca	gaactgtggg	aaatgaaattc
29401	ctgtgattta	aacccctggg	tctatggcat	tttgttatac	cagcccaagc	ggaatgagac
29461	aaaggcttta	tggggaagag	actgagccac	ctcccacagg	ccagactgca	gccttcttct
29521	tatcctgggg	ctgagtggcc	gagatgaggg	cagcaccaga	tataggaacg	aaagaaagcc
29581	ctggccaggc	gcagtggctc	acRcctgcaa	tcccagcact	ttgggaggcc	aaggcggtgg
29641	atcacctgaa	gtcagaagtt	caagaccagc	ctggccaaca	cggtgaaacc	ctctctctac
29701	taaaaaataca	aaattagcca	ggcatgggtg	cgagcgcttg	taatcccagc	tgctcaggag
29761	gctgaggcag	gagaatcact	tgaacccagg	aggcgagggt	tgcagtgagc	cgagatcgcg
29821	tcatgtgact	ccagcctggg	agacagagca	agactctgtc	taaaaaaaaa	aaaaaaaaaaa
29881	aaagaatgaa	agaaagtgtc	atttttcccc	acgtttgagc	ctaagcgtgt	aaatgttgac
29941	ctcacgctcc	acacagaccc	tcagcagacg	cacgggcccc	tcacagctc	tcaatgggac
30001	ttctcaaaact	aaactgaccc	tggcctcttg	tgacagccac	cacacccctg	gacccagaaat
30061	ggcctgggag	agcctgacct	ggtctctcct	tccccactgg	gctgaggctg	gcacccagca
30121	gacccagcag	aaccactgcc	tcaccatgaa	ttgtggctct	gagtagatca	gagccagccc
30181	tcttgccagg	aggcgagag	ggacactgat	atgctacgag	accagctcca	ggatcacccg
30241	cctcgaggga	aaggcaggga	accagactag	atattagtg	gaatctgccc	actggggcct
30301	ggaagagcgt	ggcatgtcat	gagacctgag	aaacagaaca	gggtagctca	gcaaagcagg
30361	gatggcgctc	ctggcagagg	aggcagccca	tgcaaaccctc	ccagggttgg	aagcagcgag
30421	gcaccttcca	agcacagagg	ggatgtggaa	gggatgcccc	agcagaggag	gtgggagggg
30481	agcactgagg	ccagaccacg	aagtccaggag	aaggattcca	aatcctctct	aagagcagtg
30541	ggaggacatg	gaggaaggat	gatggcgggg	catttaagaa	gcttactcca	ggtgccttac
30601	ggaaagtaga	tggaaggggg	caagatgtgg	gaatgtgggg	ttcagctttg	gggatgagtg
30661	cagagaaggg	acggatgtga	gggatgatgt	ggattggaat	aggggtcatc	tgccccgcct
30721	gagcagggag	aggctggggc	tctggaaggg	cattcgggca	gctccccacg	ggcgaggggg
30781	ggagatgcgc	ggctggccct	ctcctggctc	ctggccaggc	cctgccccct	aggacgcaag
30841	atcttccac	tctcctagcc	ctacccctgg	ggcgaagggt	ggcagctctc	atctcaagaa
30901	ggaatcttgt	gtccttctga	acattgagca	gatctaattc	tgctccctga	tttgggacag
30961	gaggtgacca	accactcaa	aagccccgtc	gctgggcatg	gaactccctg	ggatagaccg
31021	tcagctcac	attctgcttg	gaggtgggg	cgaggcagct	gagccctggg	gtgtcaggct
31081	gctaggcggt	ggcccttggc	tggggagctc	actgtgggc	tgctccgccag	cagccagctc
31141	ccacctcctg	cgctgggtgc	cctgacctg	gcctctggaa	gcctcagatg	gttcgagat
31201	taactggggc	tccctggctg	ctgaggagct	aSagagggca	tgcttgcaaa	tgctcccagc
31261	aggtagccag	ggagaaggag	gccaggccgc	agccaaggcc	ccctcagccc	tcctccctg
31321	gtagccttca	actgtggtgc	gctctccctg	atcctctccc	caggaaacct	ctgtcccttc
31381	ctcccttcca	caccagcaca	gcacatagtt	ccccaggcat	gcccactctc	tgggaggcag
31441	gcctgagcaa	ggcgtgggtg	gggcttgcga	tggtccctg	gctcctctgg	gtggctctg
31501	ggcYtgggat	cctgctccgg	agccctcaac	ttggccccct	cccaggccag	gtctcagcag



```

31561 cccctcccag cagtcaggaa ccccatgggg cactgtggcc cagagaccac tgaccactgg
31621 cctgggtggt tctcctgccc ctgctgtgct tatggcaaag tccaaccac agtccggggc
31681 cgggctcctt ctgctgtcca gcctctcagt ccccgtaggt taatatgata catcaggaac
31741 agacatgggg ggcaggcagg gtttgtgggg catgtgagag gcaggggtgc tccacagtgt
31801 ccttgctctc cgtgggcctg aggagtgagc cccgtcctgc cccctctacc atcccagggc
31861 tctacctgct tttcaggagg ctgcaaaatt gccaaaattt cctgcccagc cctggccaga
31921 ggctgagagg gccacactgg gcatgggaac cactgggagt aatcagtgtg atccccgaca
31981 aaacactggc gtctcagta aggacacagc ggcaacattt ctgcagagca gagagtctgac
32041 tgatgtggca aaaagagcaa ctgaagacag aaggcgacct ccacatgcct gctcctgca
32101 catccccctc ctccccggc cccaccact ttctctgcca gacaggagac aggcggtagg
32161 gctcccaaac ggtccccctc gtgtcctccc cccccccca cccacacaca ccattcttca
32221 ctacgatcac caagcggaat agatctcacc ttccccctcc actgtctaaa accctcaacg
32281 gctctccagt gccttcggaa gcagggccca gccctcgcct tggcattcgg agccctctcc
32341 acgctggccc tggctggcct ctccatgcgt cacttcagat acctcccac atggcaccct
32401 tgcctctgac agcccagcca gccttcccca gagcgggctg cctctcctg ctgacctgtg
32461 tttgggtcag gcctggaatc ctactctcc ctgtgcagcc cctcaattct ttagatatct
32521 caggagcatt tgaaaaaaa gactcatcct aaaaaatata tctcaaagga aagaggcagt
32581 gaSctccttg cttaaacctt tctacccca aggcattgaga ccgcagttcc tggcgtgacc
32641 tccaggccag gagggtctgg cctgctgag ccttctcggg ctgcccctgg cctgctccct
32701 gtgctccagc ggctagcccc gctgctcctt gcacaggcag cctgtctcca cctcaggaac
32761 tttggacttg ccgtgcaact gctcccagca gtgYctagcc catagtaggt gctcagtaaa
32821 catttgctga atgaatggaR tgaatcactt tgaagatatg gaataacagg aaatgggcaa
32881 aatgtatgac ccagcgtagg cctgtgacct ctctcatggg gaggccccca gctaaggggc
32941 tccccctccc tccccccagc agcctgggg gcctcctccg catccacct acccaaggcc
33001 tgcgtgtgct ggggagaagc agctctcct tcccaggttg cagRgggagg ggaagcccag
33061 ggacaatgtc cttagtact gtctgagaaa ttgatcctgg aatctttaaa ccttgggggg
33121 aaatgttcaa agccgcagcc atctctgtgM ggagagaggc caagggctct ctgttttggg
33181 gaaacacatt gtaatgcgcc ttttgcttat gcaaagaYgc agccaagttt ctgttctgga
33241 atagacaatg aggaaggagc ggcctctgcat ggcaatgtaa agtttgctc ctggtggca
33301 gagttccacc caggggatgg tgagccctgg gagggtggcct gcggtgctg ggagagggtg
33361 actttgtcct ctgccaagggt gggggagggg tgcagaaagg ggggaaccgg tgcaggctcc
33421 ctactgggga ggcctactgg cctgccagcc acagaggaga caggcagggg ccttggaaca
33481 gcagccacc ccaacctcag gaaagtcccc atRtctctcg ctcagctgag ctttgggcca
33541 ccagaggcag cctgtggccc acctggcaaa accatgtcca ggagcaagcc aggcctgaaa
33601 agtctgtcag aggcctgggg gaggggggcc tgctgctctg tggcttctct gggacccccct
33661 gaaataagtg aggagaaggc tgttctctct gtgccccctt ggaggccagg cttgaccac
33721 caggatggag gcagagaaga gagtgcagcc aggagcaggg gacagagggtc tgctggcctc
33781 cgtggaggaa gagttaggc attgcatggt gcctaaacgg ggctgccaga gcttgggctt
33841 atcatgagaa tcaacgttcc cattgcagat ctctctcacc ctggctcctc cagcaggtgc
33901 ttcagaggga atgctagaag aaacgtgact ctgtctcaaa atatcaaact acctgaaNc
33961 ggaaatggcc aggatcagaa ttactaggct cagttcatac atcaaaaagt gaggtggtca
34021 gagtccccct ctccaatacc cagcgcagga acactctgcc caaggccacc cctcactcc
34081 tcaggcgacg tctcctgtgc ctgtccacac tctgttgata cctctgcaga tgccctcagg
34141 tggggacggg gctcctccac ttccaggcca gctcatcaat gcacagagcc cagcagggctc
34201 tcaagagatg tgctgcaccc ctcagtgagg cactccagca ggcacggtga gaagcaggt
34261 cgcattacaa accttcccgg gtttgtacta acgacggcaa gaggctgccc ctccccctct
34321 cctgaatttg gggggttagc agagtgtggg ggggatattc cactctgag gtccagattt
34381 atgggtgcca gcatgggctc tgccacaatc ctgctgcgtg acttggggcc agttacttaa
34441 cctgtctggg cctcagtttc tttatctata aaatggagat aataaYagtc cctgtgtctc
34501 aggggtgttt cagggtgaag caacaatgaa aattatcacg tgctgccatg tgaagagatc
34561 accaaacagg ctttgtgtga gcaataaagc tttttaatca cctgggtgca ggtgggctga
34621 gtccaaaaag agagtacgca aagggtgggt ggactaccat tcgttggtat aggtttggga
34681 tagacgggtg agttaggagc aatttttttt tttttttgag acggagtctc tctctgttgc
34741 ccaggttggg gtgcagtggc atgatctcag ctactgcca gctccacctc ccaggttcat
34801 gccattctct tgcctcagcc tcccagtag ctggggctac aggcgcccgc cagtgcgctt
34861 ggtttttttt tttttttttt tttttttttt agtagaggca gggtttcaca atgttagcca
34921 ggatggtctt gatctcctga ctccatgac tgcccgcctc ggcctcccaa agtgcaggga
34981 ttacaggcgt gagccaccgc gcccggccag gagcaatttt ttgtgggctg gggacggatt
35041 ttacaaagta cattctcaag ggcgggaagaa tattacaaaa tatcttctta aggtaggggg
35101 ggacaatatt acaaagtatc ttcttaagga tgggggtggg gaagaatatt actaagatc
35161 ttcttaagtt ggggggagag aatattacaa agtatcttct caagggtggg gaggtgtat
35221 catacaaagt agattcaca gggcgggtcc ggcggggggc ggtggcaata tcacaaagta
35281 cattacccca agggcgagga ggggtgtatt ttacaaaatt caattgattg atcagctagg
35341 gtggggcagg aacagatcac aatgggtgaa tgccatcagt taaggcagga actatctatt

```

```

35401   ttcactttctt   ttgtggatct   tcagttgttt   caggccatct   ggatgtatat   gtgcagggtca
35461   caggggatat   gatggcttag   ctgtgggtca   gaggcctgac   attcctgtct   tcttatatta
35521   ataagaaaaa   caaaatgaaa   tagtgggtgaa   gtgttggggt   ggtgaaaatt   ttgggggtg
35581   atatggagag   ataatgggtg   atgtcttctca   gggctgtctc   gagcaggatt   aggggcagca
35641   tggaaaccta   gagtgggaga   gattaagctg   aaggaagatt   ttggggtaac   aggtgatatt
35701   atgggggtgt   tagaaggagc   atttgtcgta   tagaatgatt   ggtgatggcc   tggatgcagt
35761   ttggtatgaa   ttgagaaact   aaacggaaga   cacagcgtcc   gaataaaaag   agaaaaacag
35821   gtattaaagg   actaagaatt   gggaggaccc   aggacatcca   attagagagt   gcccaagggg
35881   gttcagcata   attatttgc   tggttggcga   gtttttgggc   tctatccttg   agtttatgtt
35941   gtcatacacc   aggcagact   gatttaggt   aaaacaacac   ctagggtgatt   caaaagcttt
36001   attgctcaca   caaagcctgg   tgggtgtctc   ttcacacgga   catgctgtac   aaaagtga
36061   tccccagcac   ctagtacagg   ctgaatgaaa   ctgtcatcaa   attacaggag   ggtccaacag
36121   cccaagctgt   ttccctggga   ccccatgcac   agaattggtc   acctgggctt   cgcccttaca
36181   cccaccctgg   ctctcctcct   ggccatcacc   atcgggagag   gtctgagtga   acagcttgca
36241   catgcccctg   tgttttccag   agctggagg   gctgtgtgtc   ccaggcacRt   ccgcgagtgt
36301   gaccagccc   tggagctctc   cccgtggct   gcgtgcctt   tggggaaaca   acagccatgc
36361   caccacttag   ggactgacct   tcccaccgc   tgaccggatg   cctcatcaga   aagggaaatg
36421   ctgcagaacg   gggaaaggacc   agtgtgtctg   aggtgtcagg   acgcagaagg   acacggccga
36481   ggctcaggc   cagagcccag   tgtccccag   gggcctctcg   gccactgcga   gccactgaca
36541   gggaaaggcc   gagaaaagtc   agggattctR   attcttgcta   gcagtggctg   caggacggcc
36601   ctgcctccct   ggaactctgc   tgagctgtc   ggctgtccct   ggccacagag   ccctgggggg
36661   ctgagtggg   agtttcttt   caaaggggg   ccctaagtgg   ccagggtctg   tgccaccccc
36721   ggccctggc   tccccccaca   cgccgtctc   agaggctgtg   tggcccggcc   catcccgtga
36781   ctgagctcg   tcacctgggt   ccgtttgagc   accaccatgt   aatttgagaa   caatgcattt
36841   ttattcccc   acacatcata   ttctgcccgt   gctctgtcac   cagtgtgacc   taaggaaaga
36901   gggaaggaaa   atgccctat   gggaggagag   agctgtggca   tcaggggaca   ggaacgggtg
36961   gggggagggg   gcagcaagca   gctgccggg   tggggggcgg   gctgtgagc   ccacaccagg
37021   cttggagccg   gagccctatc   cctaccgtg   ccccgccct   agggtgagat   gcagatagtc
37081   cccggggttc   tgggctccac   tcccactga   gatgagacc   tctgctgagg   gtggaggcca
37141   acccatatg   tggccaagca   ggtggttaga   atggagagg   atctcatgtt   aatgaatggg
37201   cctcaggcag   cacatggatg   cccctcgctc   cagaagggg   tgtgtgggat   gcagcagctc
37261   ctgctggcga   gggccagggc   gtgggtagct   ggcactggct   ggcagaggct   gtgagatgga
37321   caccatcaa   ggcatagggc   ttggactgg   cagggttcca   gcatgcctaa   gtggcccagg
37381   cgtcagaccc   gaccacctca   gagcaaagct   gaggaggact   caggccagcc   tcctactcc
37441   ccaggcatcc   atgcatgatg   aggaagaagg   gagaggccct   ggcagagcag   gtatgaaccc
37501   ctgagagact   gagaaggcat   gacagagggc   tgcaggccat   caccctgcac   tbtgcccctc
37561   gaccctgggc   ggggccccca   aggcactcag   agcgcagggc   aagagttaagg   tgaaggctgc
37621   aagaagggtc   ctgcattaca   agtccactgc   aagggtcgtc   tgctgccagc   gtatccctcc
37681   ctgaaagcgc   agtcatgcag   ggtgctttcc   aacaccacga   gcagatgttc   cagcgcctct
37741   tgggcttggc   gggaaacggc   tttgcccttc   tgtgcagcag   gtcacaatcc   ccctgcacca
37801   ccagagtctg   tttatgtggc   ttagctccct   tgccagactc   tggcagatgg   aaagatagat
37861   gaaatcctg   ccaccgctac   acaggaaaacc   cagaaccaa   ggaagagggg   ccgaattatt
37921   ggaagcaca   gggtcctgta   tcactttgc   tcaggatcag   aagggccagc   atgggcagcc
37981   cccaggccag   gctccctgcc   aggcgcgggt   gggagggtctt   gcaggctgca   gccaggtttc
38041   cttcctcctg   tctggctgca   gggaaagttt   ccttgctctg   aagtttcctc   actctgtgaa
38101   cattcccacg   gggcacctta   tgcctgcctg   ggtgctgtcc   catcctggta   acttctctgg
38161   accatcgtga   tctccagtgc   ccttcgtttg   ttagctgtct   gtccgtctcc   actaggcagg
38221   gactgttctg   cattcttgc   gatgcWct   ggggtgctcag   tactcccgtg   tttcccagg
38281   cttcaggacc   aaggccaagc   tcagccgcac   gcgccttctc   ccaccccagc   gggctcccag
38341   ggtggccgag   ggctggcacc   tcttttgac   aaggggctcc   tctgtgtgga   atgatgccct
38401   cccactttcc   acttgggaga   caccaactct   gtgacaaaa   ccagctcaag   gatcacccca
38461   ctgtgactat   ctccacKacc   aggtggagac   agctgaccgt   ctcttttggc   tgccaagRcc
38521   ctgtcctcRc   tgtgccaagg   caagcagggt   gccccatga   gaaaaaagct   tgtccaggac
38581   gggagctggc   tggggctctt   ccagacatc   atgggcccag   tgaacaactg   ctgcctcgc
38641   tggctggctc   tgatcttctc   cgccctYgcc   gagtacttcc   ccgcgtgcga   tatctttccc
38701   ttcccacacc   ctgcgagtga   gtgttatccc   cgctgtgcca   gtgagggaagc   cgaggcacag
38761   agagattgag   aggtcacat   aagttgggac   gggagttagt   acagaggctc   tgaggagcaa
38821   tttgtgcaac   ctgctgcaac   tagaaactgt   atcccgggac   ctggccattc   agacgtcagc
38881   atggacgatg   gagacacagc   ctctgtcata   gaagcgaagg   cataaggaca   cccagccacg
38941   aggagtcca   caccacaga   tagagttcat   ttttatggga   gtctctgaga   tcacagtgc
39001   tttatgggaa   ttcatttagt   cacttatcta   ttcaaaagag   atttactgag   cactactac
39061   gtgccagggc   ttttccaggt   actcagtga   aaatcagggt   ggcagaccct   ctctgggtcc
39121   gatacaagct   cactttgtg   ggagtcctta   accctgctga   gacacaccac   caaggagacc
39181   cctggactgg   ggaggaagac   ccaggattgc   ttgactgcaa   ctcagctcca   ggggcagcct

```

```

39241 gggagggggcc tcagggagtg agagcccacg ccaggtcagc ccgacagccc ggcgtggcac
39301 atcatcgccct ctgttcttgt ggttcagctg tcacctcctt gctggagcct gcacaagaca
39361 agctccgagg ctgctccaag ctgtcatccg ggtgtgtgga tggctggccc tctctgtcca
39421 cagggactct aactcccat ctgacaacca tcagttcccc gcgggaggcc aaggacagga
39481 tgctgggagg ctccacatc agctatcagg aggtctgcgg aggaggagg agggaggagg
39541 tggggagatg tccactgggt gggagggtgc tgcacatttc aggaggaagc agaggtgaga
39601 tccacccaca agacccctgt tcgagtgttc cccacctgct caccatcact ggtctgtcgg
39661 cctccccagc cctgctggaa catgtgtctc ctgctgtctg gatcccgtgc tcaccacgac
39721 atccccagca cctagggtcg ctctatcccg gtgactaag ggagtgaagt ggcaggcggg
39781 Ygggcctcag aacaggagga cacagtagat ctatgcagaa gccaaaggca agcaagaggc
39841 agagcggatg gcctgYgggg agaggcgggt gtggccaggg gactccaaat gctgtgggga
39901 ggtcagtgag gcctgagatg agaaccYgcg gagcatgcta ggggatttgg gagctctcga
39961 tcttcaggcc agacagtagg ggaaggaccag cacaggggcc accttcccac aggtatgtga
40021 ctccatctgc cgtgctctgg ccatgtctct ctcagtctca gtgacctgat gcaggcaagg
40081 gaacctcaca gggccactgg ggcaaggtag tgaggtcaca ggtgtgcctc ctgtcacagg
40141 gccagcacgt gcaaaaaccag caaatgtgag tccccctcct cagggtggga ggggtgtttc
40201 caaggataga aagcataagc cccaaacctc ctaagagaga ggagtctatc tttgtagacc
40261 acaaatcaac ctcgctccc cctctccaag cccacaagct ccagcagggc ccaatctcgg
40321 cactgcgatt cacgatgcgg ccaggccttt gcgcgcgctg tgccaccacc aggagtctc
40381 attagaggca actgcgggtg cccttgtccg aggagccttc cctgaccagc cgccccaggc
40441 accttccct gatgaaccaa ggtgtaatca cgtgtcctga tcccgctccc catactcctg
40501 tgctcYgagc tccccagtg caggatgcg gaggccagga gtccaggagg tgccgaacat
40561 gactgaggac agctccaat cccagctatg tectcccaca cccgtgctct actgagcact
40621 ttctctggtg tcaactcgag tgggtgctcc aacagccccg ggagttggca ctgaccagca
40681 cagggcgcag caggacaggc agtcgccttc tctctctag gctggaaaga acaactcaggc
40741 caccacaca ccagacttgt cctgcagccg tgttctgctt cacctgtgca gtgtggcttc
40801 ttaatttcaa tctgtggccc acattgaaaa ataattagat tccccaaaaa aagtagattt
40861 ccagcttctc attggaagagc gagctggcag cagtggctct ctattccta cagaacaca
40921 agcgagtggg accacgagga gctctgcagg ggggcgtggc ctctccggtc gcccgagtcc
40981 ccaccactcc caacagcctt cagccacctg ggcctctcgg gcctggctac tgtgagcacg
41041 tgagggcact gccctcttct gagccttgcg tgcagcctcg cacacaccac tggccctcat
41101 ggctgtgtct cctggtggca cacagggcca ggctgtaagg ccagcgggac aagcgtgagg
41161 aacagaaaca tggccagcaa cacccccagg gccacagcac cactagcaga cctgaaccc
41221 cgcttgcttc cccccactta agggaccctg gtctctgtgc cagtttcaga gatgttcttc
41281 tctttcccag gaaccagagc cagcgactcg tgaactatga ctggggcgca cgccttccgc
41341 ctggctgggt tgcatctctg cggccagttt acgagatttc caccacgacc agctgaaaac
41401 actcctgtcc gaccactgcc ttccccacc ccaccatgcg gtctcagcac tcgtgaaagt
41461 tggtagaata ttgagcccat atgcttccg attgtttgca aaccttggca aggtctggt
41521 aagtcaggc caatttctgg attaaattag aaatgaaaag aggaaaaata tctgaatcct
41581 ggcatagaga gaaatcattc attcaaaaaa aacggctcgta gaacatactg tacagagaat
41641 gttctcgtgc tgggaggggt caaaggttat agccgcgact gaaggcgag aagatccttg
41701 gattccacgt tctccatagg cctaggagcg tgtctgcct cagcagaaga gatgctttgg
41761 gggctcaggc attctccag aacataagca gtttaacacag ctgggagaga ctattctgca
41821 tgcccggaaa cctgctcttc ctgctctgcc gttctacagc ggttaagaSc agttcattct
41881 atctttaaa aagaatcccc gaagtgcct caccggctct ttgagctgat aaggctagtg
41941 tgggggtggg ctgggggtctg agcagccac tgtgtcagat acttccctaa tacgtacacc
42001 cacgtgcact gatgtagtga aaggcacatg ctcactttgg aggcctgggg ccatcacagc
42061 tcaaggcagg cactaagcaa gtacatttct gcacagcctg gtagcaaaaa ttcctttagc
42121 acctcatagt gaatgttctc agctttgtgg gccatatggt ttctgttgca actaccaaac
42181 ccactgttgt agcacaaaaa tggccacagg caattcataa atcaataggc gtggctgtgt
42241 tccaataaaa ctttatttac aaaaacaggt gaagggccag gtggccctgg ggccagagtt
42301 tgccaacaca gctttagaat ctccagttta ccttgacaaa ttcacgtgaa ttttgctttc
42361 atgcaagtaa tacgtccaaa tttactttaa tgtaaaaact atgtttttaa atgcttacca
42421 tccaaaaaag gggccacttc agggggagat gYcttacatc tgtctgggtg cctctggcag
42481 gcctgagagt acgtgttctt ggtagcaata ggttgagcgc tggccttgct gcaggctggg
42541 tatcaccac tgcaatggtc tgctcctgt ctacaaggag tgccgtggga ccaggggctc
42601 acaaagacag ttcaaggctc ttgtagatcc tgatatttaa gcaacacatt gaggccttag
42661 gcttctcggg aagagttaga ttgaaattag cgcttttagt aaatgcttga agccacaag
42721 gaagcttcag ccttggggcc aagctttata tttcttgaag tcttggcaca gatagatttt
42781 acttttaaaa atgtatctgt agccaggcgt ggtggctcgt gcctgtgatc ccagctactc
42841 aggaagctga gacagcagga tcacttaagc ccaggtgatt gagtccacca tgcagcatag
42901 catgaacctg tctctaaaaa aaatctttaa aaaatgagtc aggcattggt gtgtgtgctt
42961 gtagtcccag ctgcttggga ggtgaagtt tgaggctgca gtgagctatg attgagcac
43021 tacgctccag cctgggtgac tgagttagac cctgtctcaa aaaagaaaaa aatgcatata

```

```

43081 taaacatatt acttccacaa cggctctgat ctcccaaggc attcatccct gcctcccatt
43141 tacctgagcc tgcagaagag tccctgtctg tacctgggga cgaaacttga ggtgccacac
43201 acatctgagc tgccctctag catggcagcc tgcctgtca ggcccaagcc cactctttag
43261 ggggtgcccag gccacacac atggccacag catctggttt gtgagcagag actagcccag
43321 gcgctggcag agcaagtfff tgaaaaatgtg agaaatacca aagccccagg gtcagtatca
43381 ggcatgttca gcaatgtcac ctgtgtccca gcctggcagc acccagtagt tcagggagta
43441 tgtttcccaa gcaggaaagg tcaaacatga cagtggcttc ccaatgccct cgggggaggg
43501 gggcgggect gggcagagga gagccctta tccctccggc tgggcccagg ggaggcagca
43561 acgcaagggtg gaggctgcag gctgagccag agtagaaaagc tgggaggagc agggacagaa
43621 tgaaggcaga gggaatagaa gtcactctgc cagtgcaggc gctggggggc cagcaaagct
43681 ggatgggcca cacatctgcc cacggctctc actccatcct cctagcctca cagtgtttgg
43741 aaggggagacc cctctgccct tagacggggg aatgtgggca tccccaaggg gcctcagcat
43801 ctctgaccag gctaagggcc gctgagcctc aggagaccag actcatccca ggggtggccg
43861 cagccaggag ggctacagct gcgtgtgtct aaactccagt gtaggccggg agcagtggct
43921 catgctgtg atcccagcac tttgggaggc tgaggcgggt ggatcacctg aggtcaagag
43981 ttctagacca gcctggccaa catgggtgaaa ccccgctctc actaaaaata taaaaattag
44041 ccaggcatgg cggcacgtgc cagtagtccc aggtacttgg gaggtgagg caggagaatt
44101 gcttgaacct gagaggtggc ggttgcaagt agccgagatc atgccattgc actccagccc
44161 ggcgcagaca gcaagactct atctcaaggga aaaaaaaaaa aaaacaaaac tactgcagca
44221 gagcctctgt ggacaagggc ccgccccctg cctgccacag gctcagcaca caccagtgcc
44281 cctcttagca cggctggctt cgccctgggc tcagaggtgt gtggacatgc ccaagcagag
44341 gctggggaaa ggccacagag aacgggagtc tgaggtggca tccagggggc acatgcaggg
44401 cagcagggtgt tccctcaggc tatggggtct gagcgcaggc ctcaggtttc tgggctcgcc
44461 ctaggggacag gctgacttca tagatgacca gggcctatcc tcagaaggcc ccgaccttgg
44521 ggcttgaccc tctgcagctc tgtcttgaag ttcttcgtca tctcaccctt gcatgtgtgt
44581 tttgtaagtg aaggctcctg gggcaatgca gcatggctcc atctccccag gacaggcttt
44641 cagctgcccc cagcaggggc aggatcaggc atgtattccc cacggcatct cagggcaggg
44701 gtggtgggtg ccattccgcc cagggctggc aggaaagggg catgtctggg ggtgggtgg
44761 ccagagcccg tcttccggca gaggtagRcc tcccgtagt gacacctgca tagagatggg
44821 attggccgtg gccatacgct ctgggctgga gcagtgcacc agtaggaagg gggcgctgct
44881 ggctctactt ccccagacgc cctggccagg gcacagccaa tgtgttcttc caggggtggc
44941 aggaagctct cagaccccag tgggtgtgccc ctctctgctg ggggtgcact cactctgcta
45001 ggacccccta cctgaggagc ctccccctgg caggattcca ctctgtgaaa tccctctctt
45061 ttcctccag cctccctgag tctggctgtg cctgtcataa acttctggtc agcctgaggc
45121 acccatcagg gatcagccac aaaatacaca tttgtttaat ttcagtgacc ccacatagaa
45181 gctaaaatgct ctaatatattg catttaaaat agccatcgca caatttaaa atgaataaga
45241 aattcatgct aataattcaa gggtccaatt tttgttccact tagaacaata ttaaaaagca
45301 aataaaaaga acagcgagtt gagagacaga gagacagaga ctaaggaaga aaggagtggt
45361 cttaaatgac cctttgtatg gcaccttccc ctgctctttg aacaagcgat cccactttt
45421 cattgtgctg tggggcctgg gagttactac tggtctgccc tttaggccag tgttccccac
45481 tggtaatatc ctgacagatt ttgcagcagc cagatattga ggggaaataa gaatgcattg
45541 catgtaatac attggctcct ttgaattctc tatctccccg ccacaccttt ctttctttt
45601 tttttttttt ttttttttga gagagaattct cactctgtcg cccagactgg agtgagtggt
45661 cgtcacctca gctcaccgca acctccgect cccaggttca agccattctc ctgctcagc
45721 ctcccagta gctgggatta caggcacgtg ccaccacact cggcttattt ttgtattttt
45781 agtagagatg aggttttgtc atgttgccca ggctcgtctc gaactcttga cctcaagtga
45841 tccgcccgc ttggcctccc aaagtcctag gattacaggt gtgagccact gcactaagcc
45901 tatctttctt cttactttgg gagaaattca ggtggagggc tgctatggct aaaccaaga
45961 acagtcagcg gttgggtagg aactggagct tgggaagtac agctgccaa agggcagaa
46021 actcagaaga ttaacttctg gggagggcag gaggcattMa acacatcacc caaggagaa
46081 gcttttcccc tgggaaggga aagactgcag gatctttccc tgtggtcaca cacacaaaa
46141 tgtaggacat caaaaactat cccaggcgga ataaaaacaa aggtcaacaa caggactggg
46201 tgaaatattt cccacacatg tgacagacaa aaagtttatc tacttaagta cacagactct
46261 tccagatgaat aaaacacccc caaatcccag cagaatcaag ggcattgaata ggccaagcaa
46321 aaaaataaaa taatctctgt ttacaaatac acgtatattt ttaaatgcag gctcattaac
46381 atcaaggaaa taacaactta caacaaaaaa acttatctcc gttgtaatta aattaccact
46441 gctcaaaagg actcagtggt gtgaggaggt ggggagacag tcatttgtct aatgaagcgg
46501 gaatagatga aaccataaca gagaacattc cagtgtagaa agcttaaaaa atgggcatat
46561 ctgttgccca gagacttcgc ttccagcaac tcagcttaag gaaataagta agcacgagcg
46621 atagattcaa ctctatgaac gtgatacagt tagtggcatg tctgacctct aatggagaaa
46681 taagttttgc tatgtgaata tgtcatataa ataataagtt aggtaaaaag tgtgtgacag
46741 tcatatgggt cgatgttatg caatcattaa aacaacgtag tagacaatta actgctggga
46801 gaagacattc acatacctac aggaatttca gctaatttta catagYaaaa aaaatgtccc
46861 ggcacgggtg ctcacgctgc taatctccag actttggaag gccaaagggtg gtggtcacct

```

```

46921 gaggtcagga gttcaagacc agcctggcca acatgggtgaa accactgtct ttactaaaaa
46981 tacaaaaaaa aaaaaaaa aaaattagac ggccgtggtg gtgcatgcct gtaatcccag
47041 ctactcagga ggctgaggca ggagaatcgt ttgaactcgg gaggaagagg ttgcagttag
47101 ccaagatcac accactgtac tccagcctgg gcgacagagt gagactctgt ctcaacaata
47161 ataaataaat aaaaataaat atctgtgtta aatatggact atttcaataa accaaaaaac
47221 atgcttttct agatgtaagg aggaagataa gaaaactgga cccagcaggg gctgctggag
47281 tgggcagggtg ggtggttgcc taaaccaaag agcccccttc tgtgcccagg tggcctcatg
47341 aagtgtctcca tttccaagcc gcttcccttc tggcaagtct gagatgaaaa gggcagcttc
47401 atctttccca tagaaaagg aatatcacgt cttgtcttta acatagacag gaagcagccc
47461 ctccagtgat ttccactgct cccggagcca caggcggctc cctccacct ctggaagact
47521 gcggcgcccg gccttgcccg ccccgacttc cctacactgt gccccctgct tgacaggcca
47581 ggctgcctcc attcttccag ccggccccat tectgctca gggccttggc gccgcctgce
47641 cctcccccac aacctggaac aggcgatccc tatcctacct ctcccacat ccttccgttg
47701 tcgtgagcct cWgctcccg gatatctcca gtgcaccagc cctgagctg tagaaagtgg
47761 cctgcttcc cctacccttc tctcttctcc atctcctaac tattttattt ctttcattgc
47821 atttgtcatt atttgaatg atcctgtttg tttttcactt cagtatcgtc tgtgcttccc
47881 ttcccagaat gcaaggtagc aggggttacgt cctgtcatgc tcacagctgc accccttgcc
47941 tgacatataa tagacgcccc ataaaacacc gcccccgat gaacggcagc caggggcacc
48001 agacagagga ggatggctat gccgagaatc caccatctc cagcagctga acacaataca
48061 ccgcacaccg tgtttcctta ctacacacc ggcgcggggt gggtcgggag ggcccgagcc
48121 tatgttgtag ctgtacctac tgggacacat ggcttccaa gtttccgtgg caggggagac
48181 agagatggag gggactccaa gctcttacca gccttgggtc acaggtgacc caggtcactt
48241 cctttcacac aggccactag gcaaaactgg ccatgcagct ccacaccagc tacaagggag
48301 actggggagg ggagcacagg ggtgatcgat gaggtgaac tgcctgtgcc acccctggag
48361 cccatatccc agcccaggct ctgccacaaa gagatcccaa agggagctct gcaggcagaa
48421 gtcaccccaa atgatgagg aaactgaggc agtgaggggc tgccagggtg gagctgtgcc
48481 tgactcagc ctcgccacc ctcactctcc cctcagagcc tgcattccct gggccccacc
48541 catcatggcc cccaccagct cctctcccca gggctcgagc gcctttccca gggcagggc
48601 tgtgccacca ttctctgcat ccagttcagt gggcagctca ttgcccaggg ctgcccacca
48661 gaccccgggc tctggatgtg ccgttgccca gccccacctg ctctacagct ccagacaaac
48721 acctcactgg gagcccagtc ccagccggct gctgccacgt ctgtccttcc atgtctttgc
48781 tttcgctcca gccacRggga gccctgtgac cagctgggac ttacaagagg agtgtgacaa
48841 gccactggcc aggccttaga gggccacttg gggttgtggt tcgaggcccc catagcacc
48901 aagtccagag acagggcagc gtaggcagat cctctgtgca caggcccaac cctgccctgt
48961 tcagcacagc agctgccagc cacactgcgg cttccatcca ggtgaaaatt aatccaaaca
49021 aagtactctt cgggggtggct accgtaaaaa agaaaaaaa gacaagaggt ggtgacaatt
49081 tggagagact agaatccctg tacactgttg gtggggggca aaatggcaca gccactgtgg
49141 gaaacagcgt gaaaactcct caaaaactgt aagaaagaat taccctgtgt cccgcaattc
49201 tacttatggg catatagtca agagaacgga aaacagagtc tggaaacacc acgttcacgg
49261 cagcactatt cacagtagcc catgaggggg ccaccggagt gcccatggat ggctggataa
49321 acaaaatgtg gccgatccac ccacacacgg gatactggtc ctccctgaaa aggaaggaca
49381 ttcttacaca tgctgcaatg tggatggccc cgaggatgtc atgctaagtg aaataaagcca
49441 Sacacaaaag gacaattatt gttgatccca cttatgcgag gtacccagag tagtcaaatt
49501 ctacagaaaca gaaagtggag tgggtgggtgc caggagtcgg ggggagggga ggaaagggtc
49561 gaaactcaat gggcacagtt tcaattttac aagatgaaga gWtctgtgga cgggtgtggc
49621 gatggtggcc cagccacatg aatgtgtcta cgccaccgaa tccgcagctt aaaaaatggtg
49681 aagacgggtga cttatgttct gtatatttca ccgcaattaa aaXtttttaa attaaaattc
49741 tcttggaaaca aataagaagt taatttaaaa attcagttct tcaaggacac cagccacact
49801 tcaagtgtct cagagccccg gtggctactg cccagaaaag ggcgagacga gcatctcccc
49861 agcaccgcgg gaagctctgt cgcacagcag ggctctggaa ggtcaccctt tctcagaagc
49921 tcccacttgc tcaatgcccc tctcccccaa gaccaccaa ggttcgaca tagcttagat
49981 ctgctcctga gactctccgg gcatgcagag gctctcagt aggacgcagg gcattcatcg
50041 tcagggcaaY gccatgctcc aggcacggcc gcgccttctc tccctcgccc acgctgcgct
50101 cttgcagtg gactggggc tcagatcaga ggcagagacc ccccgctgac cctcgccgc
50161 cggctctcca Ygccccggga ctctctgcca cgccccgcgg cctctgcagg acacccccac
50221 tgtgcccattg acaagcacac tgtcaacacg tttgtgtaat gaatcagtgc tcaactctgt
50281 gcaacagggg acaacagaca cggcccaaatt actttctcaa cttctgagac ttgtctccta
50341 aaacggtttc tctgttgtag taaaacttca ctcaatttaa gaatttttaa atttctcatg
50401 cttttctaga aacgaacctt cccattattc aaaatagggg gtacggctgg atcactaggg
50461 tttgaataaaa gttggaYggc aaaagtgaga gaacgttcta gacatcctcc aaccagctgt
50521 tgtgacagag agaccctggg caccaggaga aaggttgatc tgctctgccc agcctcatga
50581 aggactcacc acgggccccca catttggtca gcgcccctgg tgggctatca gaagtgaccc
50641 ctcttgagct gtcactaaac agtgccactg tcatgataac aaaaccacat tttatgtgga
50701 aactgaaaac gtgttcaccc tactgagaga ctgaaaagca cattcactga cattatctca

```

50761	tgtaatcaca	tcaacccccca	cgatgaaaac	actggcccat	tttacagatg	agaaaacYaa
50821	ggctcagagg	acRggagcag	cgttcccaac	gacgcgcctg	gaaaaggcga	ggcctggagt
50881	YRccatcgga	ctgtaattcc	aagcccagtc	ctttccccgc	tgggttgga	ggcgacataa
50941	tgctagggga	agaactcata	ttgtgaaggc	agccaagatt	ttatacctgc	cagtctcaca
51001	gacctagcat	gaggacttcc	caagataatt	tagcatggca	agtggccagc	atggggcccg
51061	cacagtcaca	tagtaggggc	tcgataatgt	agccattatg	gttttattgt	tattattatt
51121	attattatta	ttattatfff	tgagacggag	tctcgctctg	tcactgagge	tagagtgcag
51181	tgggtgcgac	tcagctcact	gcaagctccg	cctcccaggt	tcacgccatt	ctcctgcctc
51241	agcctcctga	gtagctggga	ctacaggcgc	ccgccaccac	gcccggctaa	ttttttgtat
51301	tttttagtaga	gacggggttt	caccgtggtc	tcgatctcct	gacctcgtga	tcgcccacc
51361	tcggcctccc	aaagtgtctg	gattaYaggc	gtgagccacc	gcacccggcc	gttattatga
51421	ttattgcatt	aaaggtacgt	ttcaaaaata	agatgagggt	aaagtgtgaca	acgctctccc
51481	aaataaaaacg	aattctgccc	atgcagcgga	cttcggcccc	ctgtacggat	aagRctgggc
51541	tccagagggc	tagctgactc	acccaagacc	ggtcaagagg	aaagtgagca	tgaagggtgg
51601	gtctctgtcc	agtgttcagc	tcactcacat	ccctcctctg	ccctgaggta	tcttttcacc
51661	agggccctgg	ggactctgac	ccatgtgtca	gcttggtctc	tggaggggct	caccctgggt
51721	ccgaaaagca	gagctgggca	gcagatccctg	agcagacctg	cacaccctgc	tgcttgcttc
51781	tagcccgccc	ccagaggacc	tcattctacc	acagggtgtg	acacgcagca	cggggcttcta
51841	gaacttccca	cctgcccggc	ttccacacca	cctccagccc	tccttaagcc	ccggcttctc
51901	cccacaaata	ccccctgcc	cctaggggcc	cactggcagc	atctcacatt	ccaggccgtg
51961	ccctggtect	ccttctcacc	cccttctttt	ccctgcagga	ctcagccaag	ccagtctgct
52021	cccctgggag	cctgcatttt	gaagactgga	cgccccctcat	ccccacagtg	gctctgacca
52081	catgaggcag	agagaggaaa	gctctgcaca	gcctccaccc	acacacggtt	cagcatcctg
52141	gcacctcctg	cttaggttgt	tgggggtatg	gaatgagggg	gagggtggtc	tttaattctc
52201	tagtgttagg	cagaccatgg	gcagggtgcc	agaagacgaa	ggtggcacat	agacaccctt
52261	ttcaatagcc	cagactcaca	gggagaaaagg	aattccctcc	ccaattcctc	ccactgYgag
52321	ctttaaggag	gggcccctgg	ttgggtgtctg	gcccctgcctc	taatgcaYgt	gaatctctac
52381	tttaaacaga	aaagggcctt	ctgtttggag	cgagaagaag	gcagggctat	ccagctgggg
52441	ctgaataacg	tcactttatg	tcttactatt	atattagctt	ttcaggtaaa	ataagggttg
52501	ctttttaaat	gatttttagt	ttaaaagtgg	caaacatcaa	gaaaacatat	ttttcaaattg
52561	acggggacatt	gaatgacatt	gtcatttcaa	cgactggagt	ttgcacaaac	ccgccttagc
52621	ctgtgaagag	cctcaccaag	ttaccatcac	ccacgatgtg	cagacgagag	aactgaggtc
52681	agggatggtc	aacttgtctc	acctggggac	ccagcaagcc	tccttgagac	tgtgtcactc
52741	acctccggct	aacaaggaaa	tcggcagagc	ctcaggggac	ctggcccagc	tgggcctctc
52801	tggcctgccc	agcccctgg	tgaccctccc	cgccccaccc	tgccccctca	gccacccac
52861	tcacaatgtt	catgagtgtc	agcaggtaact	tctgtacgtg	ctccttcccc	tggaaactgca
52921	ccacagagtc	atccacgccc	agcaggacct	cgaagtgtta	gtcatcgctc	gcagcatgcc
52981	tgcgctccct	ccgcctcgag	ctgtttggcg	gctcctctag	gaYgcccagg	gcgcggtgga
53041	ggctgtccag	gctgtccagg	gaggccccctg	caaggagagg	acaccgtctt	cagcggcagg
53101	gcaaaccac	ccggacacaa	acctaccagg	cgccagcttg	gccatcctct	cattttaaac
53161	tcatagcacc	ctgaaccgag	ggtattatct	cattttacag	aggaaagcga	ggtgcagagg
53221	aagatggcag	agacttgcca	agctgtctca	gacacagcag	ggcccagaca	ggcacctctg
53281	tgacttccag	gaccgtctgt	tcgccagccc	ctgggtggcaa	ccccagccac	agcaggggcc
53341	acagtggcct	ccacaYgatg	gatcagacct	gctgagactg	tgagagacc	agcccagcat
53401	caccgtctgc	cttatggagc	ccagggtctg	cattcagcca	tggagagggtc	gggaacctca
53461	gcccgcactg	cctgacgtctg	gagtgggcag	agccacccct	tgcccacact	gcccgatgct
53521	ggagtgggta	gagccacctc	ttctcccccg	aagggtgtgc	aggtgccaaa	ggactgggtg
53581	cccagggtgc	atcgccagct	gttatgtactg	gggaaaccga	ggcccatcct	gggctgtgca
53641	tgttccccag	tgtctccttg	ggccccctccc	agaactttct	gctaagtcct	ccactgaagc
53701	gaggccccac	atctctgggc	cagggtcccg	gccccctccc	cagggtgaacg	gtaccgcctc
53761	tggcctgatt	gaattgtgca	gagaccaaga	tgaacccctc	tgtggcagct	ccgtgagct
53821	cggccctggg	caggcactga	cctcctgccc	ccagggtgtg	cccctggcca	cttggggctg
53881	cMatcctggg	gagtctgccc	cggatctcgt	cctgccttc	tgtgcagtgc	tgtctctgct
53941	gtgttctgtg	tctgcgggc	ctggagacca	ccctggacca	ggcttcagtt	cccagccctc
54001	cactgcccgt	ctgaaggcca	cagcaaggct	ggccagggtg	acctgcagca	tcgatggcag
54061	cccagacca	cactgcccac	cgccccccaa	gccccagtg	cctggaagcc	cccttcccct
54121	tgactggtc	tctgtgtcta	agccagggtc	cagccccagg	gtctccatgg	cattccccag
54181	gcaggctcctc	agagactcct	ggcccgggac	agacccagg	ctgcagggtca	tcgtccacatc
54241	actgacccac	agggcggtgg	ccaggccccg	aggccccctg	agctgaggga	gaccagcgtc
54301	cctgctgcac	cacctctcta	caaagagtct	tccagaacaa	cacttttcac	ccccaacgcc
54361	tctctcctgc	tgccctcttc	aggcagggac	aggcactccc	tcctgtgtgag	ggtggccag
54421	gacaggccct	ggaggcagaa	caaggccctc	cggggcgga	cggccactcg	gacagtgagg
54481	cagcctggta	ccaggccctc	atcccggcac	acgcttccctc	acccagtgct	gacagcaggc
54541	cactcgggtg	acRtgggccat	ttccgtttta	cagatgggga	gacccatggc	atcccagcat

54601	ggccgactct	cggccagaag	ctctctcctt	gcaccctgca	gcctggctca	gacaagacag
54661	actgaacatg	aagtttcagg	cacaaacctg	cagcccggct	gtgggcaaga	ccaagacaga
54721	cagtggaggt	gcgagggcga	gggtgagggc	tgggaaagag	aagaggagta	agtggggcca
54781	gggtctgagca	gctgctgggg	ttccagtacc	ttcgggggtc	cttgtgccac	ccccaggtcc
54841	acagtctctgt	gtccccctcg	cgcacacaca	catgcacaca	cacacacatg	cacacacatg
54901	tacacacaca	caagcacaca	catgcacaca	catgtacaca	cacaggcaca	cacacatgca
54961	tgtacacagc	aaggactttc	cagggccact	ccagggaacg	gggtggggcc	ggcaggcctg
55021	cgggcagggc	tgtggggagc	cgaagcagtc	acttagctgc	ctggatgact	caggggccatc
55081	catgcccgttc	ccagaggaag	cggcagcaac	tcaagcctga	cagggggggt	gagggaatcgc
55141	cactctatctt	ttgtacatctt	atttctttgc	atttccaacg	aacagagaca	gctcaccttc
55201	agaatcacct	tgggaaggag	agtgatggca	atttccatag	gaaaactcct	cctccctggg
55261	gaacaggccg	ggcgttaccc	tctgcccttg	tcaaactagg	ggtcacacac	gggcgggtgcc
55321	ctcggcaaaa	ccagaccaca	gaggcactcg	ttgaggtcac	gtgggtgccat	ctgggtgata
55381	gtatcttga	actgggagat	ttcatgtttt	aaaaatctag	aacatccagc	ttctcttgag
55441	gaatlaagag	atggggccac	cccaggggta	gcccKggctc	gtgaattgag	gggccacacc
55501	tggagacaaa	acacacaggg	cttggctcaY	caggccacct	gcctgccact	gtcaatgtca
55561	ttctggcacc	agaaggtatc	acagtatgtc	ctcctggccc	taaccccaga	gtccccatgc
55621	agggaccctc	ccagacatct	cccgtaggag	ggcatgggtg	ccgcacatcgg	acagactctgg
55681	gctggggcac	Rtcttgaggc	tgcttctgcc	gagcaggcag	agtttttaac	taaactgtgt
55741	gttgcaagcc	cattaaaaac	agcaaaggtt	tctaaagatt	tagaatctga	atttcacctg
55801	aaatgggaatt	attatgtgtg	ccatcattat	tttaatcgtg	ataatttKgg	ggtgaaattt
55861	ggaaattatg	ttctagagSa	tccctccacc	caggaccagt	gcctagagag	acggcccaag
55921	gcccaaagcc	aaagtgggtg	aggagaaggaa	gacaggggcc	tggggccacct	ggcagggtgac
55981	agccacggcc	tgcttctggc	ccctggaaagc	ctcagtggtg	ctcaggctgg	acaggacaag
56041	gggagaagcc	ctaagagcag	ggggatctgc	aatgccctca	gacctggggc	aacaggagct
56101	ctggcccagc	ccccctcctg	tcctcctctg	tcccaacaag	acccggcagt	caggctgaag
56161	ccccagcttg	tcaatcacga	gtgccacctc	agttttcccc	acaataagat	tgaccttgac
56221	ctgtccaatc	acgagtcca	cctcagtttt	tcccacattg	ggcttaaccc	aggccatgaa
56281	gagactctga	ccacatgctg	gcccctgcag	acctgYtccc	caaataagagc	cttgctatag
56341	ggagctctgg	ccagtgggccc	tggcggtggc	ctgagtcgtc	ttcctagtga	tcctcttgaa
56401	gtcaaccttg	gccttggggc	tgcccacctg	cactgggtgc	ctacctgggtc	taacagcccc
56461	tgttccactt	ggctccctcca	aggcagaaga	gggtatattg	ctgctctccc	ctcctccccct
56521	gggagctccc	tgtgtagccc	agccctgctg	agctgtgggtg	actgggagct	cacacagcct
56581	ctgctgtatc	cctcgccact	caccccgctc	ggatgagtga	tcagcgccacc	catctcctgc
56641	agaggcagga	ggatgacact	gcccttggcc	ttccttttct	ctgtgtgcag	ccttccccac
56701	aagaaggcac	tgggccccct	gggggtgctg	attgaccaag	ccccagaagt	cataccagag
56761	tgggctgcag	ctaccttgaa	gggaggtgtc	tgtggcagga	tcccatcttt	taaataggta
56821	acacagagct	ttcgagtccc	tggcattccc	tgtgtcagg	gccccttttg	gatggcagct
56881	ggagatctct	ttgctcacat	cgtggccaaa	Kaatgacact	gtgggaagca	ctccagagcc
56941	atcctgtttt	agagtctcta	gaatggaaca	aaaagactct	atcgaagagt	caatcaataa
57001	acaggagagag	gatggaatga	ggggaggcag	cagttctgtc	caggggaattc	catcagggga
57061	acgggtgggt	cacagttagcc	acgggctcat	ccacagtgcc	acagtgcaca	gggcacgtgg
57121	gcaagggtcg	caactgccag	catcctaggc	acaccccgag	tctccaggca	ggggtgtgtg
57181	cacagccctt	cacagtggca	gctccatacc	cccagcagcc	atccaatccc	ctcgcaatagg
57241	tggaggcggtg	tcaccatcag	catatcgtcc	tatttcagca	tcagggaaac	aaaccaaggga
57301	gcaggaaaat	tacatgactY	ggacaagcca	caccaccaga	gaggaggagt	gggcacgcgc
57361	ccaggcctct	gagtccaacg	ttgtccacat	Rcccggagcc	cttctgtact	tcaccagcac
57421	tgagggtgta	tttgcgRgaa	tgagaattgt	agaaacaagt	atgattagac	tagagaaSag
57481	aaaacctggg	actgagtcR	caacctcggt	agcatcgtgg	gatgggtcgg	gctctgaggg
57541	tgggtacagt	gggcaggtgt	gggatctgtg	ggcaggtgca	gtgggcaggc	acgggctctg
57601	tgggcagggtg	tgggctctga	aggcaggtgc	agttagcagg	cgtgggcccct	gaaggcagggt
57661	gcagtgagca	gggtgtgggc	ctgagggtgg	gttcagtggg	cacgtgtggg	ccctgaggggc
57721	gggtgcagtg	ggcaggcatg	ggctctgtgg	gcaggtgtgg	gctctgaagg	caggtgcagt
57781	gggcaggcgt	gggcccctgag	gggtgggttca	gtgggcacgt	gtgggcccctg	agggcgggtg
57841	cagtgggcag	gcacgggctt	tgtgggcagg	tgtgggctcc	gaaggcagggt	gcagtgagca
57901	ggcgtgggtc	ctgagggcag	gtgcagtggg	caggcgtggg	ccctgagggt	gggttcagtg
57961	ggcaggtgtg	ggcccctgag	gcgggtgcag	tgggcagggtg	tgggcccctga	gggcgggtgc
58021	agtgggcagg	tgtgggcccct	gagggcggtg	gcagtgggca	ggtgtggggc	ctgagggcg
58081	gtgcagtggg	caggtgtggg	cctgaggcg	ggtgcagtgg	gcaggtgtgg	gctgagggca
58141	agtgcagtgg	gcaggtgtgg	gctctgaggt	caggcagggg	ctttgcagtg	ctaaggggtg
58201	ctgggtgttc	gctgtcccct	gtgcgaagct	cagtcctatg	ccttttctgg	gctggcctct
58261	tcacctgcga	tgcccctttt	gctgctgccg	ctgtgctggc	caaatttagg	ccccctgaaa
58321	gacccagaag	gaaggagcc	cgctgcctcc	catgggactc	agagctcttg	aaagacagga
58381	gcatgttctg	cccccatctg	cctcctgcac	agccacactg	cgtgccacta	ggggacgaca



```

58441 gggacagggg ccggggcagg gatgcttaaa acctgaggac cacactgttc agtgacacgg
58501 ccgactgctg tcccctctcc atgggatggc cgccccctca ttgctctctt ccttcagggtg
58561 tctgctgaaa tgtcagctcc taaggagaaa ctttccttat ttcattccaa attgaaatca
58621 gccctcccct gatttctccc gacagcagcc cgtgctcttc ctctctgtga cactccactg
58681 caggcagggtc ttgtctcatt tgtgaatttg cctgggcatt aactgacttt cctctcaga
58741 tagagaaccc tgggggatgg agaccatggg gtctgggtgt cctcagtgtc cccagtgcc
58801 cacaggtaac cgaggggctg tgaggggccc ggagccatgg cagagagggt ctctgggaac
58861 agaggggcca gcagaggcga gctgggggct tctggggagt tgggaactgg gatagaagg
58921 gaggcagggtg cagcaggggc ggaggggcca cactgggggt ggggagccga gaacaggagc
58981 aaaagtgtgg tggcagtgg agaagcatga gatggaacaa aatccagcta caaattactt
59041 gcatatggca taaccaagtc caaatgacca ggaggagtta aaacaagtgg ggcaaagaaa
59101 cgtaccatgt Mattgccaac caaggcattt aaggagataa actgataaaa caaacaactc
59161 atggaaaaaa atacaaaaga tacgaaactt tatgctccta aagcttcaac ctggtgcaag
59221 gaaggagaga caaggccaca gtgataggag cactgtaata catttattcc agacaagtct
59281 gatcacagcat taggtaggga tagaaactat ttgaattaat agggaaacct aaYgttcaaa
59341 cccaacagcc agaacacgtg acctgtctca gcacacacgg gagaagtccc acattttacc
59401 acattaggac acaactcaaa aactcaaaag ccgtgcccac agtcaccatt actgagacac
59461 attaggacac aactcaaaag ccgtgcccac agtcaccatt actgagacac attaggacac
59521 aactcaaaaa tcgtgctcac agtcaccatt actgagacac attaggacac aactcaaaaa
59581 tcgtgcccac agtcaccatt actgagacac attaggacac aactcaaaag ccgtgcccac
59641 aatcaccatt actgagacaa aatcagagat aacaataaaa ccttaaaccc taacatctg
59701 aaattaaaag caccRttcta ataggcctag actaaataga aaatatacat tccactctg
59761 gctcaagtag aaaacctgaa gaaaccaata acctcagaag aaaYtgaaag agcaactcaa
59821 gaaatgcttc catagaagag tttaacccaa Yccacaagga acagatactc ctcatctcac
59881 ataaaaaat gcaaagatca agacttcttg aacataggaa taaccgcRa gaaagcagag
59941 tggaaaaaag agcttatcag ctataacagc aaaactccca agtatctaag aacaaacaaa
60001 aactaatgaa tgttcaacga ctacacaaa agtagtataa tatagtctaa gRtctccaaa
60061 aacacaaaag aagacctcag taaatgaagc tacatggaca taatcctgga ctgggaagtg
60121 gtctctctcc aaatgaatca agaaatttgg cataatctcc atcaataaat tgttaaaaa
60181 attgttaaaa ttattctccg tcaataatc tccgtcgaat aattgttaaa attattctcY
60241 gtcaataaat ctccgtcgaa taattgttaa aataattgtt aaaattattc tccgccaat
60301 aatctccatc aaataattgt taaaacccaa caaagtgatt ccaacattca tccggaagMc
60361 cacaccagac agaaaaataa Ygaaaaagaa tgacaatgag gaggcatatt ttttcaaat
60421 taatgcagtg tattataaaa taatattaat gaaaaYgtag ttgtgtcaat gtcatgaaag
60481 aaacagatca atggagctga ataaaacttc caaaaacagg ctgggtgcag tggctcacgc
60541 ctgtaattcc agtactttgg gaggccgagg cgggtggatc acctgagggt gggagttaga
60601 gaccagcctg accaatatgg agaaacccca tctctactaa aaatacaaaa ttagctgggc
60661 atggtggtgc atgcctgtaa tcccagctac tcgggaggct gaggcaggag aatcgctga
60721 accSgggagg cggaggctgt ggtaagctga gatcacgcca ttgcaactca gcctgagcaa
60781 caagagtga actccgtctc aaaaacacaa aacaaaacaa aaaaacttcc aaaaacagat
60841 ccaagaatat gtggggttcc caggttaaaa tggcRgattg aacacagata tctgatttgg
60901 ctctcttcta gaatcccat aaaaagttcag taaagaaatt taaaaaaaat taaagtccac
60961 aagcctcYaa ggaataggSag agcaggagg ggtacatcaa gaaagctgaa gtggaggct
61021 ggaaacagat gggcaagtgg taacagactt ggtacatcaa gaaagctgaa tcccaggctg
61081 gccatgaaga cagccacatt ctacaaagtc tcccggagaa taagaagggg gaaaatcagg
61141 ttacaggaga ggtcatagac aaagaattgg aatcagaatg actcagagtt ctcaaaagca
61201 gccattcgatg caggaaaaca gtggaggaaa gccttcagaa ttctgaagga aaatgattgt
61261 ccaactggaa ttctgcatcc agacacattc cccatcaagt atgaaagtgg aataaggaca
61321 tttctagaca tgtcagttct caaaagattt atgtgacaag cactcctctc aagaagctac
61381 tgaaaggtgt gcaccatcaa aaacaaaaaa ggcaacaaag aaagagtaag acacagatat
61441 agaaaacgga agcccaacac agatagagaa aacgggagcc caacacagaa gacatggcct
61501 gggaggccca tggagggtgg cgagaagtga cctgagggtg atggctctgc accagaYgtg
61561 gagagcgact gagccatact gagcagcggg acccacgggg tgaactgtct tcaccaagac
61621 cccacctcca tcaggccatc tgtagcccgg ggacaaaatg cccaggagag cctggcctag
61681 gttcttggca gtacttggct tggccatgtg ccgatgaggg gcgctccctt cctctgctga
61741 ggacgcacaa cattcacaga agtttccagc tttcccccca ccgagagctg gaggcaggcc
61801 tggcgagctc tagaagcagg acgtacaggt cagccccctc cctctgaccg tatttccgct
61861 gcaggtccag cgcgtgtcc ccacctgggc cctaccaggt gcctgactgt gctgactcca
61921 agacttcttc gagaccttgc tcacgacttc cccaggaggg gccttgtcag ttcccgagga
61981 agcagaagtc agaccagac ctggagactg atctgtcac ctccgggagc ctccatccag
62041 cgggggcagt gccgccctct ccttacactg ccagcgtgtg ccagctctcc Rctttaatat
62101 taatcgtctt tgatattcat ttgcattgaa ttataaatag ttgctcaagt agcaatacag
62161 tgaataaaat tagtaaatga tacaggtagc aaacgggtgt aggaaacggc ggagtgagg
62221 caggagggga caacctggca gtctggcact actgagggga tcatgaggaa gagcaccatg

```



```

62281 cggctcatat cctaaagata caggatgtct atcggatagc ccagtgcctt ccttggactg
62341 acccttggcc catggggccat tgttcggccc catggactgg cttgcagatg ggcaaaactg
62401 cacaggctctg taaggaggca atggggagct gcccttgca gcattgcttg ggtggagtga
62461 ggggagcttg aaacaacacg ggggtcatca gtgggtgaga ggagaagcaa atgtgggtggg
62521 ggcattccac aaagtgtctac caaggttaga agtttgatat ctacatagat ggatggatct
62581 taaaaacaga gtgccaagtg caaaaggtaa gaaacagaat gagacacaat tcagagcgac
62641 ttacRtaaat tatggatata cagacgcaag acagcaacac tcatgttgca aaaatgtaca
62701 ctaaggaaaag gatacggttt gatcacctga aatggttgct cactgcagga gaaaatcaga
62761 aggaagagag caggctgggg agataaatgg cttttgtttt tttctttgtt ttaatgatgt
62821 taggacaact ggttagaggg aggaaaagg caaattttga cctcatgcct taagcaaaaa
62881 cagattccaa gtggatgaga tgatcaaaag gaagtcattg aagtgtctaga agaaaatgta
62941 gcKaggcacc tttttatcat cttgctgttg ggcgggcttt ggggttccaa cctagggctc
63001 tttcccagcc ctctccttag ctgccccatc cctgtgagc agtttcccac tctgctggac
63061 tctgctggct gaatgcaagg gcagatggca cacttgctag gggatgggtg gaggggggat
63121 gggcacactc Wctaggggat ggcgMaagg ggggatgggc acactcata ggggagagc
63181 tgagggcaga caggcacact cactaggtag gggatggcct gagggcagac ggcacactca
63241 ctaggtaggg gatggcctga gggcagacgg cacactcact gaggcacctg cacacagaca
63301 gactgtctgt gctactgcct caacctgcca accatgtgga tgggaacaag tcattccttt
63361 tccagtgtct aggagtcccc aaaacgtaag tccaatgac cgcccYgcag agccagaact
63421 gccacaStgc ccagaagagg agagggaggc tcagggaatt agactaaggc catgatgttg
63481 gcaaatggcc agggcagaat gccagccag ttcctgcaca caagcactgc caggggcctgc
63541 agggggcacc accRggctca ggaggaccaa agcctggagc accagcagcc tctcctccaa
63601 ctgaggtcct cagaacatac ccccgcacg ggaggtcaac tcatccctc catttcttac
63661 Yaaccaagag agagtgcctc tgtaggtgtg tattccctc aagataatca ttaagatgaa
63721 atatgtttat gtggaagtgt aatcaggaca cagcagaaag ccctgcttgg aagcaggcag
63781 gctcttttga ggctctggga agcagaaggc aggtcagcc ccacgggggc catcagcaca
63841 gagccccact ggtgtacttt attaaaagat gctcattagg attattacat cccagccctt
63901 tataacttta ttaactgact tgaattaat tttgcaact caacagacac agctcatcca
63961 atgccatttt tatgacctga ggtggatcgg aggggtggga ggggctcctg gggccggcct
64021 ggccatcagg agcgcacag gaaccgctgt gccgaatctc aggggcagtc ctgacagttc
64081 tctaaactcc caYgggatgg gctgccccgc ggacagatgg acgctccagc ccaggcctgc
64141 tcagcctcag ccRtgcttca gggagaaaaa ggaagcacta aaggagagaa ggcctctcgc
64201 ggccgtcct cagaacatac ccccgcacg cagaggacaa aggaaccccc tcagtgggtg
64261 acacagaagg tccctggcctg cccagcact gcacaccctc acagagccac ctgggggttg
64321 acagcacggg ggcctgccct gtctggccat tacacctttt ctgctagtaa accacctcc
64381 cctatttgca ccccatatgg ctcatgggSa gccaatccca ctgcacccca gctccaggag
64441 tgcccaaggg gtctgttag gccactggaa cctgatgctg tgtgttgggtg tgcacaggaa
64501 gactcccgta ctttctctc tgctgggaaa atgtgactct gggRctgcta agttactcct
64561 tccccagga gcagcagagg cgagagatga gggcagcccc gtgtgctgct gacactgttc
64621 aggtcccaa atccagctat ggctgaagg gatgcacccc tggacatcct aattcttcag
64681 gtcaattcat ttagctctct gtctcttgca accacgagag tcctgtgtgc ccttacaac
64741 acagacacag caaatgtgc tggcctgaga gagacacat ggtttggctt tgcagccaca
64801 gctctcttga acgatgagg catcagacca ggcactctc tctgacctg tgctgcacc
64861 tcccagtgtg gggcgaggag aaggctctgca gtatttaagg tctatgacca gaaaaggggc
64921 caggctcttg tccggctccc agatacgtgt ccttgggaaa tgggtgccag cagaggtgtg
64981 aggtgtctga ggttgagccc aagccagcag ggccatgctg agtgaccagc agctgtgcga
65041 gcccctctga ggtgtcctgc agcagggtct gggcttctg tctacagccc tcacagagg
65101 tcccaaacac aacttcccc tcttcccccc agggctgggg cctcctgtct acagccccc
65161 cagaggttcc caaacacaa tttccctctg tccccccagc acccgccgaa ctaccgaggc
65221 tctctgcac ctgcccctg ccactctcca gcggccYctg cagccactc acctccaagg
65281 attcacgggg gtctcggggc ctacagaggc tggggcgggt gacaagagcc agcaagccct
65341 gcccctgctc agcatctgct taagccaggg agggctccca ggactacca cagcccccag
65401 ggctccaagg cttccctctg ctgggcaggt gggccaggga aggggagctg gctgggttaa
65461 tgtcacgggg gcaccgcagc acgcggcagc caagtgggtg gggccgatgg agaggccggc
65521 aggtctttta ttgaaaatat tcaactgtgaa cttcagaaca aaacgaagat ccccacgggt
65581 cccaaaggaa gggagaattc ccagggggtg tgttctgaac gcagaRgagg tgggggcccgg
65641 caagtggggg gctgacagcc agtgccacgg gacctatgt gccggagggt tggccagtgg
65701 ctacagagat caccagcgga aatccctgt caccagtcYg gtacatgctc cagagccct
65761 gatttggaca tcatctggaa gcctggcatt gctccccgga aggcctcagg tacctcacac
65821 tcagaagggt cagggcctgc tgccctcctc gcccgacgt gcacctccgg ctgactgcc
65881 agctgcatcg tccacctagg gagtggggcc cgcattgcag caggctccatg ccaggagcag
65941 aacctgcctc tccccctccc tcaaaaggca ctcccctc cactcctcagg gctcttgaa
66001 cctaccctc caccgtctat gccctctgt ctgcccacgc agcagcatct
66061 gcgcctcccc catctctgca ctcttcaaat ccaagtccca aggcatttct ttacttgcca

```

66121	acctcagttt	cccttgttgt	ggccttcacg	tcctttgatg	ttgagccctg	cgttcatgca
66181	gggtgggct	gggcccctgca	gggtgaggt	cctgggtccc	tgggaagagc	ttgcagccaa
66241	cacaagcggg	aggtctcaga	tgctgcccga	ccctgtgcaa	atgaggccct	ctcctacccg
66301	gccatgagcc	acgtgcaggg	actcagcacc	tgccctctgct	ccctcctcct	tcttgacccc
66361	tctaggtcag	gagtccaggg	acacagcagt	gYtggggtgc	tgcagccaag	cacaggctct
66421	tcctagctga	ttgtgggacc	tggggcagcc	ctgcacgggtg	ggccacactg	tgccaccaag
66481	ctgccttcct	aggaggaaga	gttcctcaga	tgcccaggca	ccagggaactt	cacacacctc
66541	ctgtttcttc	ttcatggtag	ccggtggagt	gggcacggga	cctctcttac	gcaaattgag
66601	gaacaagccc	tgggcccagg	caacaggcac	ccctgcgtga	ccccagcatc	tggctctgtc
66661	ccccgcattc	cgagcttcca	atggcccattg	gtgggtgattg	cacctgggtcc	agcagccaaR
66721	aggagagctga	aagtaccagg	gccctggccc	agctgaggac	cttgctgcca	tcacccacca
66781	tgacccttgg	gcacaaagac	ccccactttt	cagtcgtggc	accaagcaca	gtccatgtct
66841	gaccaggacc	tgctcaagga	gcaggagtga	ccgaggacaa	ccccggggac	caccccagcc
66901	ctcccccatg	ctgggggtgg	aagggtgggtg	aggcctgggg	cctcttgatg	ctcccatgcc
66961	ctgatgtccc	atcctccagg	acgtctccat	gcccctgcgc	ctcctgaggg	ggcagggaaa
67021	ccccatgtgg	gcgggcagca	gccaggctcc	ctggatctgc	ccagagaaga	gccccccgca
67081	gaggtgatgc	aatgctcacc	acgccccctt	cagctcaaga	ccctggaact	tgccagaccc
67141	tcagccctct	ctcccagtg	gtgtcagga	cctccctttc	ccctcaggct	gctggggccc
67201	caaaccacat	tgctctcttc	cctctgtctcc	agccccccat	ggggctgagg	agcaggccctg
67261	cccaaccttt	gtttctggtc	tctcctgtcc	ttcacgcctg	gctgtgggat	cttctctcag
67321	acctgcctct	cccacgggta	aatcccaagg	tgcagactca	gagacagcac	cgagaggccc
67381	ccagtgcagc	cgccactctg	agcccagagg	ccaggccatc	caaccctgca	gggacggggg
67441	ttcctcccat	ttgtcttctt	gtgtgaactc	cgtgggtcca	gaagcctgag	tttgggggtg
67501	gtgacaattg	gcctgggtgg	accccacaca	ctcctgtccc	atgatgcttc	ctgcccataga
67561	atatcccagc	tgcagttgag	gagggaaagg	gctcctcata	ctccccacac	agcaccacaga
67621	ccacggtgtt	gacgaggctt	aatttgatca	gccctatttc	gagaagggag	gaaatcagaa
67681	atcagagtga	gtgctgagtc	ctgaaaattt	aatccccaag	gcagcctcac	agggcctttc
67741	ctccccaaac	ggggcagggtg	gcaggagtgc	actacaagcc	cagatcccag	ggccaccctg
67801	agaaccagga	ggggaggagg	ggccaccctg	cctgcagaga	ggatggcaaa	ggcctgctt
67861	gcccgggccc	agctgccctc	aacactgcaa	gggctcagaa	gactggacca	gcagctggcc
67921	gggcagaaac	tatagaaatc	atcctaattc	tactttaaaa	ggctgggggtt	tcctcttctt
67981	gcttaaaagc	tggtgtgacc	cttccaggta	gcagccagcg	gcaggcctgt	ggctcactga
68041	gatgggctgg	gtgggtggcc	ctctggcctg	tgctcccaac	cagctgcccc	tgtgtctctg
68101	cctgccttgc	agaggtgacg	tggttgccat	cacagcatgt	catgttgtag	actgactcct
68161	gccaaagccct	gagagggcct	caagtctggc	agcagccctt	gcaggggcca	gctgcctcgg
68221	aacctccagc	atgcttctcc	caggggcagga	gcaccctccc	cagagcacc	agtctgttg
68281	ccactttctg	ccaggcaacc	cctccccctg	cttccagggtc	cagcctgact	ttcctcctgt
68341	ttgcaacagt	tctacctgc	ctgcccctgg	cccatgccc	ccaccaggac	aagtactgt
68401	ctccttctct	ggctcccaaa	accttcattt	tcatttcattt	gttcattgtt	cattactctg
68461	tcaagtattt	tgctaactac	catttgccag	ccttgctgtg	caccacctct	aggagagcgc
68521	tctggggtgt	caccatcaca	ggctggcctg	ctcagtgatc	actcctcaga	cacatgcaca
68581	cacacacatg	cacgcagtca	catacacaca	tgcacactca	cacgaatgca	ctcatacaca
68641	cacacacgaa	cgcactcaca	tacacacatg	cacactcaca	caaatgcact	catatacaca
68701	cgaatgcact	cagacactca	caaacacgca	cacacgaatg	cactcacaca	cacgatgca
68761	gtcacatata	cacatgcact	cacacgaacg	cactcacaca	cgcacactca	cacacgaatg
68821	cactcacaca	caaatggact	cacactcaca	tgcacactca	tacgaatgca	ctcacacaca
68881	tgcactcaca	ctcacacgaa	tgcactcaca	cacacgcgaa	tgcactcgca	cacgatgca
68941	ctcacgcgtg	aatgcactcg	catactcaca	tgcacactca	catgaatgca	cacacacata
69001	cactcacaga	cgcacactca	cacatgaatg	cactcacaca	cacgaatgct	cttacatgca
69061	cacaaacgca	catgcataca	ttcatcaca	gtgagtgtat	gtgagtgcac	gtgctgtat
69121	gtgagtctat	ggatgtgtgt	gcagtgcaca	cgtgtgtgca	tatgtgacat	gcacacWcag
69181	gcacacacat	gcacacacac	aaacgtgcac	atgcacacac	aggctctaca	gtgaactcct
69241	ggaagactga	ggtggccctt	gacccctctc	taagtcccca	ggtacccaca	tggggcttgg
69301	caaaatgaat	aaataaacgg	gcaaaaacg	cctatKggat	aaatggaaaa	cttagctatc
69361	cttatgctac	ttctgagcag	gtggagacag	gacagggtag	gacgaggcgc	agcgtggatg
69421	cgggcggatt	gtctgtatct	ccttggtggg	agcagttggc	cgtggagtct	acatcggttt
69481	ccttctctgg	cgtcatctac	agggagagca	tgagggcggc	gtgaggtcag	cggggcagggt
69541	ggatgtctgt	cgttactg	agccatgcgg	tgggtgcacg	ctaccacgt	cttctctct
69601	ttttcttggg	agctgtcgcc	atggctcccg	gcacaYcctg	acatccgtag	gcgccatcct
69661	gacaacacgc	tgcgggccgc	aggcctcaca	ccacagcaat	tctcactcgg	gcctccctga
69721	ccctcgacct	gtgggtccct	gctaggctgg	gcgtgtttcc	cgatcccca	tctctctgac
69781	ctagacatat	tttgttcaca	tgttgaacag	agagaaacgg	tcYgcatttc	cagaaagaaa
69841	tatactctcc	caattttcct	tggagactgc	ggccctcttg	gcttgtcttt	tgcttatcc
69901	ctcttcactg	agaaatat	tgctcccRtg	agaaaagcca	tgtgggcagg	gggatgagag

69961	ggcacactgc	ctccagccca	aggtggccag	tgggtgactag	gaggggtgcag	gtccatgttg
70021	ctgggtctta	tttttgaag	gaagtgaaga	atctggattt	gcatgtgaag	tctctgatct
70081	ctaaatactg	gcaaaaagca	attttaagaat	aatgtaggtg	ctgtgtgtgct	tgtgaaatga
70141	accaaagctg	gggtcccgagg	tctgtctgcc	acctcagctc	cacgcccaga	ccaaagctgt
70201	gtcctcaggt	ctgtctgcca	cctcggtccc	acgcccgtgag	caaccgtttc	cctgtatcct
70261	cctccgtctg	gcctggagag	gttttgagca	cagtaggttc	tggccaacca	gagtgaagtc
70321	acctgectct	gcctcacagt	ggctgaagct	tcccgggccc	cagttcacgc	cccacttggt
70381	tctccacatc	cctttcccac	cttgaagcac	agctcctgat	ggatgtgttt	tatgggtggaa
70441	tcctcacgca	ggggcctgag	gaacactccc	tggttatcca	cacggcaact	aacactcagc
70501	acgcaccagg	ctcctcctcc	ctctcatgga	ctctcaattc	agccccRcac	agccaagggc
70561	tcacccacat	catgtgcgct	tcagacaaag	gactaatttc	cttaatacat	aaagcgctct
70621	tacaaatcaa	caRgaacaa	tccagcaatc	caaaagggaa	atggacaaac	ggtattgata
70681	caaaagtgtc	gggtagagaa	gggcatgctc	cctttaaatg	atatggaagc	aggggaagtc
70741	atgggtccctg	gctaagggct	ccaccctcac	ggacctatgt	gaggacaggc	actcctgtcct
70801	ttgtgccccaa	atgtcacatt	tcccaagacc	accctggtcc	accatgtccc	catcctgtgc
70861	ctataaaaaac	ccgagaccct	aacaggcaga	cacaggcggc	tggacgtcga	gaggagcgca
70921	tgaggggagg	aacacacagg	cggctggacg	tcgagaggaa	cgcaccaacg	agcaccagca
70981	cgctgcaggc	caccgaccgg	cagaagcaga	acgacacgga	gtttggcccg	ggcagtcaga
71041	ggagagccca	ggctgcttag	cggcccgaact	ccagggaag	ccctttgcac	cccatcctcc
71101	ttctggcttc	ccccatctgc	tgagagctac	ttccactcaa	taaaacctcg	cactcattct
71161	ccaagccccac	gtgtgatctg	attcttcacg	gacatcaagg	aagaacccga	gatacagaaa
71221	gccctctgtc	cttgcgacaa	ggtagagggt	ctaactgagc	tggttaacac	aagccgccta
71281	cagacagcta	aagttaaaga	gcacatgcag	cacacgcccc	ctggggcttc	aggggtgtcga
71341	aacattcacc	cctagacact	gctgtggggc	cggagcccca	cagcctgccc	gtctgtattc
71401	tccccagag	gtttgagcag	agggacactg	aagaaggaag	ccacaccccc	atagcacgcc
71461	ctgcgaggag	gacaagagaa	cttctctcat	ttcaatatga	ggggacaatt	gtcagaaaaag
71521	aaaagacagc	agcttttctt	ttcttcttct	tctctttctt	tctttccctt	ccttcttttc
71581	ttcttctctt	ttcttcttct	ttcttcttct	ttcttcttct	ttcttcttct	tttttttttt
71641	gagacagagt	ctctgtcacc	caggctggag	tacagtggcg	tgatctcagc	tactgcaac
71701	caccgcctcc	cggattcaag	cagttttcct	gcctcagcct	tccaagcagc	tgggattaca
71761	ggtgcctgct	atcatacctg	gttaattttt	tttttttttt	gtatKtttag	tagagacagg
71821	atttcaccat	gttagccagg	ctggtttcaa	actcctgacc	tcaggtgac	catccaccct
71881	ggcctcccac	agtgttggtg	ttacgggtat	gagccaccga	gcagcttttc	agtgtatgag
71941	aagacgctcg	acctcaccct	taattaaaga	aacaacctgc	atcagattgt	ccaaataaaa
72001	atgtacggat	caggggtaca	tgctgagaga	aacggatcca	gccaacccatc	ctaggggtgt
72061	ggtcatcctt	atcaaaagtg	aaaatgcaca	aaccttttga	atccacaatt	ctagttatca
72121	aatttaatcc	tacaaatgtc	ctcatacaag	tagccaagga	aataagtgtg	aggatatcca
72181	Sggctgcatt	ataagaaacR	gtgaagaaga	gtaacgtccc	catcagtcac	ggtctggtta
72241	agcaagtgat	tcacaggatg	gaatgcctga	gcaggaggaa	tgggcagggtc	tggaggggct
72301	gctgtggacg	acgtccaaga	catgttccct	gataatcgca	aatgcagggt	agcacacggg
72361	gtagtctaca	gtgtgcagga	aaaagtccga	aaaacgaaaa	cacagtatct	acaggacaca
72421	aatcccttg	agccagagat	tctaactcta	ggatttatat	atgtgtgtgt	acaNatgtgt
72481	atatatatat	acttgtgtgt	gcacagcaca	tctgaaagg	acacagaaga	aaggtgtgca
72541	gtggctgccc	tgcagaaagg	agctgaSttc	caggacagcc	aatgggaggg	ggatgaggac
72601	agcaatgagc	atttcatatg	gctgcaatcc	ttagtccagc	caaaaagcaa	gccacgcaca
72661	tgcaggagag	gccaggactc	tcccagcagc	tccccaggg	gtcaggggca	gagggagccc
72721	ccagcctggc	ctgtccctgc	acgactctac	acttggcagc	tgaagggggc	caccctgaca
72781	gtcaaggccc	agctgtgatc	tggggtggaa	ctggctgtgg	accgatctgc	ttggcctggc
72841	tctgggtgct	ggcacggcag	tatctgggtg	gctgcagggc	ccccagctgg	ggccctttga
72901	gtgatgatga	gcctgctggg	ctgggaccac	agagtgcaga	ggaccagcat	cccctcaacc
72961	ccagatgccc	tYaggactg	cagggcccag	acacagagag	ggaacctgac	gctcatgggg
73021	agggctgaga	ggcaggcaga	gggggaaggg	tgttctctgac	cacggcccac	aggaccagca
73081	gtgcagaggg	cctgggagtg	agggaaagccc	ctgctgtgcc	aaggaggag	ggccaggaga
73141	ggagagatgg	ccgaggcccg	gaggagggca	gaagtgcaga	gagactagag	ggggcagggc
73201	agccctggct	tggggctgga	gagggagcca	ggaaggaggg	tggaaaggcag	gggaggcagc
73261	gtgtggacaa	agactgagag	ggggagtcac	agcttgggtg	cagaaagtgg	ggccagggag
73321	gattccgggt	ttctgacctg	ggcaaccggg	caccctggga	agccttccct	gcaagggacc
73381	ctcgggagga	gccgcgtggg	aggagagaaa	ggcgtgtgg	ggcgtgttgg	gctgagattt
73441	ctgtggggac	acaaggacag	acacccaYga	gacactcagg	caggagtggg	ctggagcaca
73501	gagagagggg	gggcccaggg	ccccaggtgt	gggcccctc	caggatgaat	cagctgcaga
73561	gcaccgcagc	ggcgcacaca	gatcagtgct	ctgatctgaa	aagtaggcat	aacgatgctg
73621	gccacagccc	acaagtaaga	cacatagaaa	acctgaggcc	cacagagggtg	acaggaacgg
73681	gccaggtgca	cggacagggc	tgttctctgac	tcagctcact	ccaccagaac	ccagtctggg
73741	caccagcctg	ttttcttgct	ctctccagcc	agcagggcag	aattgggggt	ggacttgccc

73801	cagcctgaag	acctccgggg	ccctggctcc	tccctgctcc	acctgccata	gccccaaacc
73861	agtgtgggtg	tgaagcggaa	ggagactcYg	cagcaaggcc	cagctcttat	gccggccact
73921	gccttgccctc	ctgcctgatg	cagtctctcg	gcagcaaggc	aggtgccctg	agacctggag
73981	gtgccccag	ccccagaacc	caccatgcgc	ggctggggccg	actgtgcctg	cccttaaggc
74041	aggcagaaca	ctaaccaaag	ccagcggaca	tccactgcct	gcggagcatc	tggggaaggc
74101	tcctgtgggg	ctccagaccc	ggggcgggtg	cttcaccacg	gcctcctcat	ccagcatcca
74161	caactctgga	gctaggaagc	ccggtgctcc	ttcccagagg	agggagccga	agctctgcca
74221	gtctgcactg	aagtctgtgg	tgtgcggtca	cctgagaaca	gctcagtcag	tctggtgga
74281	tcctgagctc	ctccccgggc	ctgtgcttga	gaggtgacag	ccagcacgtg	agcaagtgcc
74341	ctgggttctg	gtcccagcaa	agctacctcc	agctgtgcac	ctcggtcagg	cacctccagc
74401	tgtgcacctc	ggtcaggcca	ccctctctcc	gtcagaccct	cagtgtcctg	atcagtgtcc
74461	acaagtaaga	cacatagaaa	acctgaggcc	cagagaagcc	aaggaagtcc	ttacaggctt
74521	ctgtcagaga	atcggaggaa	aggaggggac	gccagagaat	ctgggcttaa	gtgacttggc
74581	cgtgatcaaa	acacaggcag	agtcacaaac	acgagacccc	gctgcccact	cccgcaggag
74641	aggtaatcc	acaaacacga	gaccccgctg	cccactccac	aaacatgaga	ccccgctgcc
74701	cactcccgac	ggagaggtaa	ttccacaaac	acgagacccc	gctgcccact	ccacaaacac
74761	gagaccccg	tgcccactcc	cgacggagag	gtaattccac	aaacacgaga	ccctgctgcc
74821	cactccacaa	acatgagacc	ccgctgccca	ttcccagcgg	agaggtaat	ccaggctcag
74881	ggcatccagc	cttccccggc	ccccaggccc	ccagctgcca	ccaagaaaca	ctctttgtag
74941	gggtcctgcc	ccactggggt	gcagaggtac	atcccccaag	gcccagcccc	gcaggaaggc
75001	cttcctctgc	agccattctt	gaaatcacct	ccaagaaaaa	agaaatgctt	cagcttccaa
75061	agatatcccc	aggcaggaaa	tggggccgag	gcagcatgaa	tgtttaatta	cgtaatttgt
75121	ccaaatccag	gcacatcat	ccctcttaag	taccttatgg	gaaatcgtaa	ttaggcgtga
75181	aatggctcat	tttcacaact	cccccttctc	gggggttctc	agggcgcatg	cccctgtcag
75241	ctggccggac	aggcctctct	tgccaaggct	taccctcctt	ggttctaatt	taattttgtg
75301	accttttctc	tgcacatcct	gaccccgctac	atccctgcgc	cttggcacac	tgacccaaag
75361	caccatagaa	ttactgagct	ataagcctct	catggaaatc	tgtttttgac	tttggttagaa
75421	aacactataa	tcaaatccct	gagagtccag	agggatgggc	ttggcctgcc	agaggaaaca
75481	tatcacccga	ttaagtgat	tgacagaggt	gggtcccggc	ggaggagggc	aggcgaggct
75541	gggcagggct	gggcgtgagg	tgggcaggca	gggtgccctc	gggggcaagg	ggggagggga
75601	tgccagagg	cagaggatgg	gcaatgtggg	attaccagaa	agtgcggtct	ggacatccca
75661	gggcctgcag	gcttctctgag	tgectacctg	gcccctggga	ctcacagggc	attgatgccg
75721	agatgctcaR	ctgtgcagag	gggtgtcagg	tctggtgggc	acttatgcgc	tccaggcctc
75781	gcttctgctg	atgccagggc	aagaccagcg	ctcaaacaca	catcaaatac	atacgcttga
75841	cgttatgccc	gggagggtcg	tgcacacagg	cacacagaga	cggtgagact	tggggatact
75901	tgggtccagt	tactgtcacc	ttgggtctca	cttctgggag	cggtgatttg	gctggcacca
75961	ctgcaactgcR	gtgtgtctgc	gtctatacag	acttccctcca	tttcacgcac	gtacctaaaa
76021	cataccacca	ttcatgtact	gtcaggcag	ctttcatattt	gtaaaacaat	agggttatca
76081	aggcctgcgc	agtggctcat	gcttctaattc	ccaRcacttt	gggaggccga	ggcaggcgga
76141	ttgcctgagg	tcaggagttc	gagaccaggc	tggccaacat	ggtgaaaccc	catttctact
76201	aaaaatacaa	aaattagccg	ggcatggtgg	tgggcgcctg	taattgcagc	tactcaggag
76261	gctgaggcag	gagaatggct	tgaacccggg	aggcggagggt	tgcaagtgagc	cgagatcggt
76321	ccactgcact	ccagcctggg	cRacagagcg	aaactcagtc	tcaaaaaaaa	aaagccgggg
76381	gggtcaccaa	aagatgagag	ggacccaggc	ctgtattgca	catccacatc	agtgcagctg
76441	taatggtaYg	cagacccccct	ccccRcatgc	acacgcattc	tgaaataagc	caggcccaga
76501	aggacaaata	cggcactcgt	ccactcacat	gaggatctgt	aatcaaacgc	atgaaagcag
76561	atgacaccat	agaggctgcc	agagctgggg	gagcgggaaa	tggggagtgt	ctgttcaatg
76621	gggtgtaaagt	tttagtgatg	ctggacgagt	aagtactaga	ggtctgctgt	acaacgttgt
76681	gcctgtagct	gaccatgtgg	cactgtgcac	ttcgaaattc	tctgagaggt	ctgagctcac
76741	gttaagtgtt	ctttccacaa	aagcaaacaa	caacagcaac	agaaatgaag	Kgacagaagg
76801	aaaccgtggg	agaagacgag	tgtctaccgc	ctcgatgtgc	tgatggcgct	ccggtgtctg
76861	cacaaatcca	cacttgctga	actgtaaggg	gccggtgacc	tctacagtcg	cctaccatgg
76921	cgagcataaa	cttgccagaga	agaggatgag	aattctgata	ataggacaaa	tcaagctta
76981	acaagtaatt	aaaggacaaa	ttgggtgcata	taacaagtcg	ttccctgtct	cgcagctgac
77041	atggatcaat	gtccaaaatt	tttatccaaa	ttgttataaa	tacacatgag	agcttctctg
77101	gtgcagtga	tagtgaagtc	agcagaggca	gtggctcacc	cctgcagctc	cagcacttta
77161	ggaggccgag	ggggtaggat	cacttgagta	ccagagttag	aggctgcagt	gagccatgat
77221	tgtgccactg	cactccagct	ggggccacag	agcagagacc	tgtctctagg	aaaaaaaaaa
77281	agaaaaaaga	aaaagaaaaa	aaaaaagagg	aaattagcca	atatcccaga	gagtgaaggc
77341	agaaccgta	tcagcttttg	ccaaaaatga	gaactccaca	gttgccctccg	cattctgtct
77401	tttaaaaagt	gtcaatttta	cattatgaaa	tagagtgcac	cttttttaag	taaaaatcgat
77461	gctccaggaa	gatattccac	gtttatctctg	gactttgggtg	ccctctcgac	acttggtttt
77521	gattaaatta	atccaaactc	cagttttgcta	agccagtgca	cctgtgtgcg	gctgttacct
77581	tgcaccaatg	ttttatcaaa	tctccgacta	tgcccaagtg	cggcgccatg	gaagccacac

```

77641 gggatccctg acagacagaa acatccatac tgctcttaag gtgggtcagg cactgttcca
77701 agcacctcgt ctatagttgc ccactgattc ttgcgagcag ctgtgtgagt tagctgtgat
77761 gcttatcttc attctgcaga tggggaaact gagaaaggct atcccaattt cacactgctc
77821 ggaagtgtca gaaccaggat atgaaccctg gcatggctac caagagtctg tgctcttgaa
77881 cgctgtgcta ggccactcct tctggggctg ttacatttta actctgacat gtcaagcaaa
77941 tcaactataag tttattgagt catttcttag ctgcttcacg tgtaaagtcg ggtgacaatg
78001 gctgcaggtc ctgccactca gacgattccc acgctgtgct gYatcaggat gaggatgagt
78061 agatcggtaa agtgccttgt agactgtgat gatcatattt ttgtctttga ttccactgcc
78121 agcacaatac cttttttttt ttaataaaagg gctgttgaac tgcaattccc atgtttgtat
78181 acattcttct tctgtttctt tgggttaaagt ttcaagattc tcttaggggt caaatgttta
78241 aatttcaaat cctttggctc atctgactaa ttacaagcat aaaatatggc atttccaaca
78301 agtctggaga atggaagcac atgaaattat gcacgaccat gagcatctgt gtcggggaggc
78361 ggaagggcgt tcacccttca aactcaagtg aaaaagggtt gccgtagtta aaggctgctg
78421 acaataaact actgttctga gaaatcacca ctgtttactg ggagaatcgt ttttcagcaa
78481 atatatcaag aaaagtaaat atcaagtata atttgcagag ttgcagataa ctccacattt
78541 tggaaataaaa aacttctgtg tctgtgtaca agtgtgcaaa acacagcaga gacatttttc
78601 tttccaatcc tgcctaataa ctgactgaac atcatcagtg ccatctccca tgactgtgac
78661 atgttatctt gtgtgtatgt gttgtatcat attatctgct ttgttaaattg ctatgcttgt
78721 taaaatgtgt caccattaaa aacaaaaaca gcccgggcat ggtgggtcac gctgtaatc
78781 ccagcacttt gaaaggctga ggggtggggga ggggtggatca cgagggtcagg agttcaagac
78841 cagcctggcc aacatggcaa aaccctgtct cctctaaaaa atacaaaaat tagccaggcg
78901 tgttggcagg cacctgtaat ccagctact cgggaggctg gaggcaggag aatcgcttga
78961 acccgagggt agaggttgcg gtgagaggag attgtaccat tgcactccag cctgagtgat
79021 agagtgcagc tccgtttcaa aacaaaaaca aatgaacaaa caaacaacaa acactgtaac
79081 aaaaaaacag aaacacacac tacactgact acactgaagc cctcaagccg ctgagcagcc
79141 atctccaggt cccaagagac caggcaccaa ggtctggggg ttggccagca tccattctcc
79201 cctttctcag agcagcagcc cagtgttatt tcctggaatt cctggaaccc tgcatgcaca
79261 cagccctccc caagccaaac atcctctccc tcacagagtg tgggatttga gccgcaggag
79321 ccagaagagg aaaaggtagt ggtgtcacta aacacctggt cacaatggtt aagacctccc
79381 tgtgtccagg accccgactg ccaggccctt ccacagcccg actccccgtg gactctgcag
79441 gtttctaat gtctctccag taactcttcc tggctcaggg cagctggact tcatgcaagc
79501 tctctgccag aagagacct ctctgccac aggcgccttc cccatcagat gccagagata
79561 tcaccagccc tcgaggctta accctgtctc tgtgggggat tctgtgtgct ctgctgaaat
79621 tcaccagggt gggcagtcct caggtgactt ctccagctct gcatggagaa ggcctcacca
79681 tgcattctcag agagaagtgg aggtgtcacc cccacaggaa gtggggccca gtccaccgtg
79741 ccctggacgc tgctcatcca ctaggactgg gcttccgtgt gcagggttcc accctcgctc
79801 tcctctctgc ctgctccaca cctgtgtgct tgacctccat gacctggag ctgctgacag
79861 gtgcccccg acccactcc agctgcacc tcccagggt ccagaccccg cccctgctg
79921 cctgcaactg cccctgggtg gtcacaaggc accttcactg caacaagtcc caaaccaacc
79981 tgtcaccttt cctacccca acccatcagc cattgtgtgt gtgcccactc catctccaca
80041 tccaaggagt caccgcatcc tggcaccttc cccatctccc cgtccatgat gagctcagcc
80101 cctccctcta cctggatgca gcagcctcct cccagacctg gccccatccg ctgctgtctc
80161 ctgccttca ctccctgtag ctgcagtcct cactgagct tgccccattt aaaatcctct
80221 gtacctcct ctcccaagc cttagcagcc cgtgcgggag tgcgagcaca ctggtcaggg
80281 cctcctgatg ggcctgacct gcctgccaac cctcccaggg acctgctcat gttcccgcaa
80341 cgcccttaca gccatgggtc cgagaaaaca ctgctcactc tgcccagaat gcctgccctg
80401 ccctttcttg cagactccta ttcagtcctc aaaaccagc tcccagctcg ttaggcccg
80461 aaagagagca agttactct tctcttagac tcaagcacc tgtgttctct tcttttaaac
80521 cactgatcat gcagaagtat aagctcacca gtttccaagt cccctaacca gacctgagc
80581 tccaggggta aaatcatgca tcagattcat ctctgggtca cctgccttgg ctgagaggag
80641 gctcatgaag gtttggcaca atgggatgaa ggaatagaca aggagcaaga atacgagctc
80701 atcctcatgc ctgagaatag catgaaatgc tcagggtgtg tgggacagaa ctgacgcYtt
80761 cacaccaatg gagccagRtg ccagcctatg tgtgtccaa gcttaccag tcccaggctg
80821 catggaatat agagggacc caaatcgaat gagcgagggt ccttaaaaga aactcacctg
80881 gcccctttgt ctgaggtcct aggggttgggt gctgccagga actctcatag gagagggct
80941 tgttcccacc tagtcccaag catctgagca gggccaacga aagtgtgtgt aggcaggcc
81001 ttgcccctgga gaggaagggc agagaggaa gaggagggaa gcgggggtgga ccacctctgc
81061 tcagaccaaa ggcagaatcc agtgcgtgat gttgccaata ggttctctc tgtgtctggg
81121 cccccatgaa ccatgtgagc cctgcacaga gctcctgcta gatcctgggg gtctcctgc
81181 actgggattg ctccaacagt gccccaccc tcagggcagtg ggttgccctc tgcatctggc
81241 caagtgcctt cccagacacc ctgcagcctc agcctcctgt gacctcat gtccacctag
81301 cctggtcctg gtcccagggt cctcctggt ctgagtgcca ctaaacagat gagaaKctcc
81361 ctcccctccc agccctgaaa tatgtcagc ctcaaagcct ctgggcacat ctgagcttcc
81421 agatgtcctg tgccttgtga gatattaacc acactctgcc tagccctctg ccatgaggctc

```

81481	agctccgctc	ccaggcaggg	ggaagagctg	ccttggggagc	aatgatgggc	atcgctccctc
81541	tgtgctgtgt	ttgggggtctt	ccctccctcc	aagtcYtgg	ctttggggccc	agaaatcact
81601	ttccctccct	cctctctgag	cacacaaagg	gcaggacttg	tcttcagggtg	cttgggggaag
81661	agggacgcca	attgttccag	ccactcccRt	aagggtgaac	caggagacta	accatcaaat
81721	tcaccaggaa	atcccaacct	ccaggccctg	actgaggcMc	taagacccaa	gcaagcaata
81781	tcaatgaaag	ccagaggtgc	ttaagggtggg	gaacaagaat	gagctgtcag	aggcctctga
81841	taaagaatcg	tttaaaaaga	ctttctttgc	tgcggctcat	gcctgtaatc	ccagcccttt
81901	ggaagtctga	ggcggtagga	tcacttgagg	tcaagaattc	aagagcagcS	tgggcaatat
81961	ggtaaaaccc	tgtctccata	cacataaaaa	aaattagcca	aacgtaatgg	ggtgcacctg
82021	cagtcccagc	tactcgggag	gctgaggtgg	gaggactgac	tgaccctgag	aggttgaggc
82081	tgcagtaagc	cgagtccatg	ctgctgcact	tcagcctggg	tgacaaagca	agacctgggt
82141	tcagaaaaaa	gaatttcccc	cttcgatgta	tgctatggac	tgaatatttg	tacccccaaa
82201	attcatatgt	tgaaacctaa	tccccagagg	gatgatttta	ggagggtggg	cctttggggg
82261	gtaattaggt	catgaggggtg	gggcctcatg	aatgggatta	gtgcccttat	caaagaagcc
82321	ccatggccag	gtgcagtggc	ttacaccagt	aatctcagca	ctttgggggg	ttgagacggg
82381	aggatagctt	gagcacagaa	gtacaagacc	agcttggtta	acacagcaag	ccctcatcac
82441	tacaaaaact	aaaaataaaa	aaaaatatgt	tgcatgtggt	ggctcacatc	tctagcccca
82501	gctacgtgag	aggctgaggg	aggaggatcg	cttgatccca	ggagttcaag	gacacagtga
82561	gMtaggatcg	tgccactgca	ctccagcccta	ggtgatgagt	gagactctgt	ctctaaacat
82621	ttttttaaat	taaaaatgta	gaaataacat	aaaagaggcc	ccagagagct	ctcttgcccc
82681	ttgcaatgca	tgaagacaaa	gtgcaagaag	gtgccatgtg	tgaaccagaa	ggttaggccct
82741	cagcagacac	aaaaatctgcc	agcactttga	tcttggaactt	cccagcctcc	agaacctaga
82801	gaaagaaata	aaactatgt	aagccactctg	gtggtggggg	tgctttttgt	ggtagcagcc
82861	tctactgact	aagacagtgt	cacttcagct	actctcaaga	aaggcttcgt	cacctggatc
82921	cagtgcgctc	agttaagcgg	tgctgcccag	ctctgggagg	ctgggggaga	tgaggcgatg
82981	tgtgattctc	aaggctcagca	tggttgaaag	tctagcaggg	tttaatgagg	acggaattaa
83041	aggagacggg	gcctgatgag	gggctctgag	cctctctgtg	tccctgtctg	ctttgtctgg
83101	aaaatgggtg	aaggaaaggag	gcactcttgg	gcactgtgtg	cctctgggtt	tccttgatcc
83161	ctggattcac	caggacacaa	aaacccagtg	cttcttgtac	tgcaagggtgc	ttgagaaagg
83221	aagagcaaac	ccacaagctg	aagagtccca	ggcagccatc	gctcccgaca	cgcacgccgc
83281	agtgtagctg	tggctcctga	gaacggcggt	gttaaaggac	agttacttaa	ccagaacttg
83341	gggtccccc	gccccgcccc	acggaggcct	tgaacatcct	acaataaaca	cacaccgaag
83401	ccatgttctt	ggagagactt	tggggaggct	gtgcactctg	gaaacagcca	agctgtctg
83461	tggcttgga	gaaacaagca	gttccatcca	ctctgggctc	acagatatct	ctcctgagct
83521	tgaaggaagc	cacagcggaa	cttggagctg	ctctagttct	ctggaacatg	cagcagggtg
83581	aaccgcagac	Rtctctccag	ggagctgagt	cccggttttc	acctgctggg	tgtccacgga
83641	ttcaccceca	gctactgtgt	agggggcagc	tcccttcagg	gactgctcct	gggagaccgY
83701	ggaggtcacc	gcagctccct	gctctgtgga	gcggccacgc	tagcagaagg	aacaggggtg
83761	gccagtgtct	tgagtatgcc	ggaggggcagt	gagtgtctgc	ggaaggggct	ggagggcaca
83821	Sggtctgcga	gtgccagggc	aSgtggtgcg	gggggagctg	tctacaaagc	cttagagcgc
83881	cagggttaggt	cccactggaa	gggaacatgc	aagcaaagac	tcKaatgaKg	tgggaagggtg
83941	agctgagggg	gtctagaaa	ggcatccagg	cgggaaggag	accccatgtg	gggttttgaa
84001	aagaaaggca	aggagggcca	Sgatggcagg	aatcaagtga	gtgagaagga	aaatggaggc
84061	cacgggggtcc	cagggcgagg	cgctgggtcg	tgtggggtct	ggttaggtggc	ccgggtctg
84121	ggtgctgact	tctgagctga	acgaagcagg	agccactgtg	ggatttccag	cagaRgacag
84181	acatgatcca	gcagaaggca	tcaaaggatc	cctgaccacg	gaccgcccc	ggacccttcc
84241	accatagtgc	aggattccct	ttacaggaaa	tgccccagta	ggcagatcta	gagaggcagg
84301	aagtggatga	atgggtgcct	ggaagggtgt	gggggaggaa	ggacgacagc	taaggggtgt
84361	gggggtctttg	gggagggaga	agatgttgtg	aatataataa	aaaccat tga	atcgtaYact
84421	ttaaatgggt	gacctgtata	gtatgtgaat	tatatctcaa	taaagctgtt	agagaaaaga
84481	atccctgctg	ctgtgttgag	gacacatggg	ggtggggagg	cagggagggtc	tgctgggggg
84541	tcccaaagtc	atcagaagag	gacaaaggct	tgacgcaagt	gacaaccgtg	gagatgggtg
84601	gaagtgatca	gggtcggggt	ggattttgaa	agaaaacct	aaaagatttc	tcaacagggt
84661	tctcatcagt	gtgagaggaa	gagaggggtc	aaccttaact	gcagagcttt	ggctgtgcaa
84721	atggaaggaa	ggatcagccg	tgacacSagc	tggggaaggc	cgagggtgga	gtgcggagag
84781	ggtgagcgcg	agttcagtga	ggggcggtgt	gagggtggtac	ttccaaattg	aggttgggcaa
84841	gtagagaaat	gccttggaac	ttggatttga	agagtgcaga	gacagatgag	gtgcagatct
84901	tgctggagtc	aggagcgtgg	gcaatatcaga	gagcatggaa	ggtggagggg	cagttggaa
84961	atgagtcttg	ggtgcaggag	agagatctgg	gctggagacg	caaactctggg	agccatttca
85021	tggtacataa	agcaacaaga	ggaatgagat	cgcccacaga	ggggtgcaga	gggagaccaa
85081	ggagcgtcca	ggacagagcc	ctgggcgc	caacccacaga	gggtgcggga	agaggaaagga
85141	tcagcaaaag	aaacggggat	gaactgtgag	tgctgtgatt	agaccaagaa	agtggggtgtc
85201	cggtaaagcca	caccaaggga	gcaggaggga	ctagcctgcc	acgtgctagg	gatggcgact
85261	gaggattggc	aggggagatgc	agcagcatga	tgacctcggg	gaggggtgct	ctggcagagt

85321	ggcacacaca	gaggcctggt	ccaagtgggg	ttaagacaca	gtgagaggag	aggaatcgaa
85381	gatggcgctct	ccacctaggg	ggtgctgctg	agagagcaaa	aagctgggag	agtaggtgac
85441	agacgagggga	agccgtcagg	tgtgaatgcg	gatagggagc	agatgttggg	ggtgttcaga
85501	ggacggaaga	tatgaaaagt	catctgggaag	gaggcaaaaca	gactgagggg	agcagctctgg
85561	ctgctagcgg	caccttgctg	gtgaacatca	tcgtaagcaa	gagcagaagc	agggctgagc
85621	tttgttcagg	ggtacccgtg	ccagcatgaa	gtaggtagaa	agtttcattt	taccaggact
85681	agggcatattg	tcagtagatg	ttacaaagca	agagtgaatc	caatgaggga	acctgcagtt
85741	taagaSaagg	aggctgggta	aggggtggag	ggaaggatca	gtggtgtgag	agaccgtgac
85801	aaggcggcag	catcaatggc	ctgcagggtt	gctggagcac	aggtcggagc	aaaaggggaa
85861	cgggaaaaata	gaaggtcatc	agaatgcaac	actgcaacgg	attatgcagg	tataggggcc
85921	gctggcaaaa	gaaggtttaa	ggtatgacca	aggccaaatc	cagtggggag	tccttttttg
85981	aaaccttgac	agctgtgggt	atggtcgacc	tcattgggaa	ttttgcaaac	ttgatccctgt
86041	gtcaaattgag	ttcccgtgac	gcgtgtcagt	gcgtgtggat	cttgcctgtt	gggggagggc
86101	agaaactctg	cagaccccat	gccttgccRc	cctcgaccct	gatacgcgat	gccaattagac
86161	acactgctgc	ggaaaccata	aacgtcacca	gattcagagc	ttgagcccac	ctctcccaga
86221	ggttttgtcc	ggtaaaaggga	agaggaaacag	ggttcaactc	tcctggcaga	actctatgga
86281	tgccgcagaa	gacaacgtta	atgagcttgg	cttaacttac	aagggatttc	cgggtYttgtg
86341	aaagaggagg	aaactcaaat	ctcaggagag	tttaattggca	ggagtgaact	tgagtgttgg
86401	tttccctaata	agatgggttt	agtgcatggg	gtgaggtaga	ccaaggccct	ctgtagaata
86461	tgctggagaa	caatgattca	agactacaaa	gaaactgaag	tgtagcggga	agataggagt
86521	gaaccttcga	gaaccggagg	ccatgccaact	gccttgtttc	gggtggaggga	acaaagcctt
86581	gtcactgtgt	tgccaggggac	agagtgcgaag	gtgcctggaa	gctcctcggg	aacgttagct
86641	gtgtccctca	tcgcaggcaa	tgggaccctg	caccagtggc	cagtacctgc	tgccccagcc
86701	agcctctgtg	agccacacac	tggcagagat	gacctcacgc	tggtggaaaa	cgaggagaca
86761	gcagatcaac	ccagccgggc	ttccacgtcc	agcaaaacga	ggcgagttcg	ggttgttatt
86821	ctgtctcttc	atgagtcagg	gctttggggc	tcactgcctt	ccaaacagag	tcaaaatcaa
86881	tttgacgttc	aaaaacgacat	gaaaattgcc	atgtcaaagg	ccattctgtg	atttcagtgt
86941	attaaaaatca	aaagagcaca	gcctttctga	agaagtttcc	tttgggcaga	attcaagttt
87001	ccctcttctt	ttacttgtgc	aatttgaatt	tatttcactt	tttcaggaag	tgaccatttt
87061	tgaggaaaaa	gtgtagtcaa	aagagataat	gcacatgaag	tttcttttta	agtctctgag
87121	atgtcgaaga	gatttacagt	gcattaaaaat	aaactatagg	caaaatagtg	tgtgtacctt
87181	tctctacgca	ttttttaaaa	ggggcaagtg	ggggaccatc	ttgttgattt	ttcacaccag
87241	tagccccctt	ccgccagcgg	gaagggcctt	gccccacca	atgtgtctct	ggtggatggg
87301	taatttcattg	ccatSctact	ccccagctgc	tggtgtgggg	gggcacatag	ccccaccctg
87361	gtactcaatg	ccctcagcac	agtgaattggt	ccaggggggtg	aacacatgac	tggagtgaga
87421	ctaatcatgg	gtcttccatg	agatctgaac	attggggagaa	aaaaatattc	ccttcctggc
87481	tcagtaagct	cacagcgtag	gcctggggct	gcgtatgtga	agagccaatc	tgagaaggaa
87541	gtctctgcag	aagagagcca	agtccttaga	ggaggagaga	cgctgagcct	tggcagcctc
87601	gtgtggggcca	cgggtgcctg	ctgtcttgct	cgcgtgtgcc	agcaaatYtc	ctttctctgg
87661	cttaatttga	tttggatctc	taattatacc	ttgaatgacc	ctgacatatc	cagtgtcttc
87721	taccttctct	attgctgtgg	ccctttaaaa	aaagatttagg	aatMatctgS	catgagaatt
87781	ccacttcttg	aatggtgtgtg	agaggagccc	catggacaca	ttccccctcag	aacaaccata
87841	actggtgaaa	agtattgaaa	agtcaccatc	aacaacaatt	taaagtctct	agaaattatc
87901	ctaagagcac	acaacaaact	aagcaacatc	atctattcaa	gaaaatcaac	caaatcttgg
87961	gaagaaaaat	cccagtaaga	acagcaagaa	tcgggtggcac	ctgagccatc	acttgctctc
88021	tcttccccgc	ttctagtctt	tccagttcag	cttatgggga	gttccactct	gggggtgtgt
88081	ggccaagaac	acagggctcc	ctcttccctc	agatcccagt	caagggaatac	agtatttcac
88141	caggatagat	gggccaccag	cgtttctcat	catctccaac	tcccagtgac	agagggtaac
88201	tccttgctga	gtgtgaccga	Raggccggag	caccttccct	caccagagcc	taactcaggg
88261	caggagctcc	gtctcaggtg	cagcagggcg	agagtgtctg	gctcctaata	accctaagtc
88321	cagcctgctt	gctagaaaaa	tattcccatg	tagggggagg	caaaccaaga	agatcagagg
88381	ctacctctg	cccaatacct	ggagtgggtg	cttcagtga	gattctggcg	gcaagaattt
88441	aaagaggagg	tgacagctct	atgagaacaa	caacctgaac	cacagaccag	cttatttact
88501	ggaaaaaaaa	taaaaaatca	ggagaaacaga	tcgttatgag	gatctcttcc	agggggagaa
88561	tgaatttcaa	agactggcct	caaaatcatc	cctacccaaa	tttaatttga	tcaaaactatg
88621	tgacaattta	tgtctcagga	cattgttgaa	aataatagag	caatcatccc	tcaatttagtg
88681	aagcctagta	gctgggtgtg	ataccaaatg	aagcagagag	tgtaacagac	agatcagaga
88741	aagggacagt	catagggagc	cctgctgaaa	tcactgtcat	cccgggatga	ccctgcacac
88801	gctcaaggct	gtgccctcca	acagctgagg	cttccactg	tgaagaatac	agactttgtg
88861	aaaacagtct	aaccaaatca	ggaaacaaat	aaatacacaa	acaactgcaa	acagcctcag
88921	agatggggga	gggggaatcag	tatgcagagc	tgccacatga	cccaaaatgt	ccagtttcac
88981	acaacagaaa	cacaaatgac	tcatttcacag	aaaaaaaggc	ggcaacagac	actgctggta
89041	agaaagccca	gatgtcagat	gtaaacagtg	tcagattttg	cagcaaaagt	ttcaaaagcag
89101	ccattgga	tatgtctttg	aaaaaaaggaa	accatgctta	tgtaaaaggag	tgatgatgac



89161	aatatctcat	caaataaaga	ctatcaataa	agagaaatta	taaacaagaa	ccaaatggaa
89221	attctggagt	tgaaaagcac	cataactgag	aggcttaaca	gtagattgga	actggcagaa
89281	gaaagaatct	tgaaaacaga	tcaatagaca	ttatgcaatt	tgaaggacac	agggaaaaaa
89341	gaatgaagaa	aagtaaacgg	aacttaagaa	aaatgtggaa	accataaagg	gcatgaacat
89401	atgcacaatg	ggagtcccag	aaagaggcaa	gtgagtataa	aggagcagaa	aagatggaag
89461	aaacagtggc	aaaaagctcc	ctaaatttgg	tgaaaacatt	aacctacgtg	tccgagaaac
89521	tcagcaaact	tcaaacggga	taaatgccaa	gagatccaca	cccagagaca	tcatagtaaa
89581	aatgccgaaa	gccaaagaca	aagagaaaat	tctgaaagca	gcaaaagaaa	aatgacttgc
89641	cacatatgtg	ggaatcccag	tgggatttcc	agccgacttc	taataatgga	agccagaagg
89701	cagcagaatg	gcatatccaa	tgtaaaaaag	ctgtcaacca	agaatctttc	tccagcagaa
89761	ctatttttca	aaattaaggt	gcaatataga	cagtcccaga	taagtaaaaa	cagagaatta
89821	tgagtagacc	cattttatca	gacatactaa	aggaagttct	tcaagctgaa	agcaagtgc
89881	ctcagttggg	aacttgaatt	cacacccaca	cacatacagt	accagtaaa	gtaatcacgc
89941	aactttaaaa	gacaatataa	atgggtat	cttctctttt	cttcccttaa	ctggcttaaa
90001	aagcaataat	atagaagaag	atgtatataa	ttgtattatt	gggcccata	aatatataaa
90061	tgtaatatat	ttgactaaca	gcacaaagca	ggtgaacaga	aacacagcta	ctggagtaaa
90121	aaaatgacac	cagatgggaa	cttgaatcca	ccgcaacaaa	tgaagagagc	cagaaatggg
90181	caacaggaag	gcaaaactgt	gatatatatg	tactctcctc	tcttttttca	gcttctttaa
90241	gcatgaaatt	atataaagca	ataattacaa	caattggttag	gtttgtagca	tatatagctg
90301	tgtggtaggc	agaactgtgg	ccctccaaag	atgttcacat	tctaattccc	agaacctatc
90361	tatctgtaag	gttatatgta	aaggaaaatt	aacattgtaa	atcagatgac	tttaaaatat
90421	tgaattatc	ctggattagc	cagatgggccc	cagtgtaatc	acaagggtcc	ataKaggggg
90481	aagggggaag	cagaagagag	aaactggagg	gatgggtgca	cgggaaggac	ttggcctgat
90541	gttgcgtggc	taaaagacga	aggacagaag	tcatgagcca	aggaatgcag	gtggcctcca
90601	gaaactggaa	agggcaagga	aacagatcct	ccccggagga	ggacccagaa	ggaacaaac
90661	cctgccaaaa	ccctgattta	aactcagtga	gactcatttc	agactctgac	ttccaaaact
90721	gcaaaaata	accattgcat	tattttaagcc	actaaaattg	tggtaat	ttaccgcagc
90781	aatcaaaaa	acacacaagt	tgcacagcaa	gaacagtaca	aagatggggg	tgggcccggg
90841	gcgggtggctc	acgcctgtaa	ttccagcact	ttgggaggcc	gaggcaggcg	gatcacgagg
90901	tcaggagatc	gagaccatcc	tggctaacac	ggtgaaaccc	cgtctctact	aaaaatacaa
90961	aaaatttagcc	gggcgtggta	gcgggcgcct	gtagtcccag	ctactcggga	ggctgaggca
91021	ggagaatggc	gtgaacccgg	gaggcggagc	ttgcagttag	ccgagatcgc	gccactgcac
91081	tccagcctgg	gcgacagagc	gagactccgt	ctcaaaaaaa	aaaaaaaaaa	aaaaaaaaaa
91141	aaaaaaaaMaM	agatgggggK	ggaagaatag	agctatataa	gagtaatggt	tctatagctc
91201	acaggaatta	agttaatata	aatctaaaga	ttctgataag	atgcatacga	caagggagag
91261	gacatcacca	aagtgcagaa	atagaggact	ccaaaattct	gccctccac	aaaaacagga
91321	ttaaagctag	taaatctcag	aatcgacttt	tttggaaact	cagaatttaa	ccagaaacct
91381	acaacaacca	gggaatgctt	aatgaagaaa	gcagctactg	aatttttggg	acagtgggtt
91441	atgggtgttt	aattttactt	cttaccatta	tcttattctc	caactcagca	gtggccatga
91501	agatggcaga	ccacattcct	ggtgtaggtt	gctgatacca	acctacacat	ccaaggagct
91561	caatgactcc	cagtttgaaa	actcaaaaga	aagaaatatg	gaccttatat	ccagagaatt
91621	ttgggtgtgg	gttttaaccc	atctgtctgg	ttgcctttgt	ttcaactgtc	tccgagcttc
91681	cccaggactg	aagtgttttc	ccagacagtg	tttgtcaaaa	gtatttaaa	gtRaatgat
91741	tagctgcaga	cacctggggc	acaggacaat	aKttgggggca	agacRcacia	taaacatatc
91801	taaaggcttg	gggaggaaga	gggtagggaa	ggagatattt	gaggaaataa	ggtttttgga
91861	aagctcttgt	gcataccaga	gaatctagga	ttctaccact	gacatgtcta	gggtgagatg
91921	catgctcaga	aaaggtctaa	gaagactcaa	aagttttcac	ctctggctga	catttagggg
91981	ctgtacaagt	aggaagtga	ggctaagaca	gagttgtaaa	cagtcctggc	aagcatcaaa
92041	ggtgtgctca	aacacagaa	caatctacaa	agacgaggag	agtatttctt	cctaccctgg
92101	cttttaagaa	agtctctgtc	aaatcactag	ctgaccacta	agctaacaaa	acagagactt
92161	tggtggccat	atataatgaa	caatgcagac	tttacaacaaa	tagtttagaa	aatcacttaa
92221	acaaacaagc	aacaacacat	aaaaaacac	aagccccggg	aagtgggggg	catctgattt
92281	ccagagtctc	cacattcaaa	atgtccagtt	tccaacaaaa	aatacaaaag	atgcaaaata
92341	acaagaaagg	atgggtccact	cacaggaaat	aaaaaagacc	ttccctgagg	aagcctagac
92401	acctcatgtg	ccagacaaag	attttaataa	aattgtctta	aatgtgttca	aagaactaaa
92461	agaaaccatg	ggcaaataac	gtaagtaaa	caagaacata	atgtatcaca	aaatagagaa
92521	tatcaataaa	gaaatagaaa	ttataaaaa	gaaccacaaa	caattttgga	gctgaaaagc
92581	ataataactg	aaataaaaa	ttcaatatag	cagttcaata	gcagatttga	gcagcaaaac
92641	aaagaatcag	caaacttaca	ttatcagcaa	tcatccaatc	tgagaagtga	aaggaaaaag
92701	aatgaggaaa	aaggaacaga	acctaagaga	tcaagtacac	caaaatacaa	aaaaatgaga
92761	gttcaagaaa	gagaatagag	aaacaaaggg	gcataaagaa	tgtttgaaaa	aacaatggct
92821	ccaaattccc	caaatttgat	aaaaagaaca	tgattctaca	catccaagga	gctcaatgac
92881	tccaagtctg	aaaactcaaa	gagttacatc	aagatacatt	gtaatcaaat	tgtcaaaaga
92941	caaagacaaa	aagagagtct	tgaaatccac	gagagaggag	cagcttgcca	catatacaag



93001	ggctctcaat	gagcttaaca	actgatttgt	catcagaaac	aataaaggcc	agaaaccaac
93061	gggatgacat	atttaagggtg	ctgaaaaaaa	tactgtcctg	cggctaataaa	ttccaacaaa
93121	actatccttc	aaaaatacaW	gaaaaatWag	acactttcag	agaaacaaaa	actgagagag
93181	tttgttttct	agttgacctg	ccctccaata	agtgttaaag	aaagtccttc	aggttgaaat
93241	aaaaggacag	aagacagtaa	ctcaaagcca	tatgatgaaa	tgaagaatac	taataaaggg
93301	aacccccaaa	tcttcgtttg	taactcttct	cttttttctc	atatgattta	gtctatagat
93361	gcgtaaagca	ataattataa	atctacatta	atagacacac	atgtatgaaa	tgtaacaata
93421	ttataataaa	gggtggggaa	gacagagctg	tataggagca	aagcattttg	tatactattg
93481	acactaagtt	gatatcaatt	caaacaaaa	tgttatgtgt	ttagataaat	agctgtaatt
93541	cccaggataa	ccaataagac	catagctttt	aaatatacag	caaaataaat	gaaaacatga
93601	atcaaaacag	tacactaaaa	aacaatgtgt	taaacacaaa	agaaggaaat	at ttgaggaa
93661	ttgaggaata	aagaatatga	catatacaaa	acaaatagca	gatagcagaa	ttcagtcctc
93721	ccttatcagt	aattatttaa	atacagatgg	actaaatgct	gcaattaaaa	ttcagagaat
93781	gggagaatgg	attaaaaaat	cccaagatat	aatccaacta	tatgcttata	cacaggggac
93841	caaat ttaga	ttcaaaatac	acaaatagtt	tgaaaaatgaa	aagatggata	aagaaatgct
93901	atgcaaatag	aaaccagaag	agacctggga	tgggcatact	aatatcgagc	aagatagact
93961	tcaagtccaa	aaatttttaca	acaaagaaca	gcattatata	ttgatgaaag	agtcaatcca
94021	ccaagaagat	ataacaatta	taaacataaa	cacacctaac	aacagaatcc	caaaatctat
94081	gaagcaaaaa	ctgacagaat	tgaagatgga	aatagaccat	tttacaataa	tagctgaaga
94141	cttcaatatt	ccactttcaa	aaatggataa	aacaaccggc	acatgatcag	taagaaaata
94201	gaggacatga	acaacactgt	aaatcaacta	gacctgacag	acatacatag	agcaccaccac
94261	tcaacacggg	cagaatgtgt	attcttctca	tgatgatga	tcatttctcca	ggatagggtca
94321	tatgttaggt	cacaaaacaa	gtctcaataa	ccttaaagag	actgaaaacca	tacaaagcat
94381	cttctccaac	cacagtggaa	taaaactaga	aatcaatgac	agaagaaacc	ctttaaaaat
94441	tcacaaatat	gtagaaattt	aacaatattc	tcttaaagaa	ccattaggtc	aggtaatatg
94501	ttacaagata	aattagaata	tataataaat	aataaatcct	gagacaaatg	aaaataaaaa
94561	cacaacatac	ccaaaactta	aaggacatac	aaaagcagtg	cccagaagaa	tattttatagc
94621	ttcaaattcc	tacattaaaa	ataataaaga	tctcaaatca	ataacctaac	ttcacacatg
94681	aaggaaaccag	aagaaaagca	acagaaactc	aaagctagca	gaaaaaagga	aataatcaag
94741	attatagtga	aggtagacaa	aatagagaat	tgaaaaacaa	taagaagact	caatgaaacc
94801	aaaagctggg	tctttgaaaag	atcaacaaaa	ctgacaaaca	tttagccaaa	ctgaccaaga
94861	taaaaagaga	gaaggtcca	attactaaaa	tcacaaacaa	aagtggggac	attactaccg
94921	accttataga	ataaaaaatg	attataagag	aatattatgt	acaa ttgtat	acgaacaaat
94981	taataaacct	agaagaaatg	gacaaattcc	tagcaacaca	gaaactacca	aaattgactc
95041	aaggagaaat	ggaaaaattg	tacagatcta	taacaagtaa	agagcttgaa	tcagtcatca
95101	aaaactccca	ataaagaaaa	acctaggacc	agatggcttt	actgggaatt	ctacctatca
95161	cttaagaat	taatattccat	ctttctatga	ggccagcatt	accctgtatac	cgaaaaccaga
95221	caaaaacatc	ataagaaaac	tacagactga	tgtctcttat	aaatataggt	gcaaaactcc
95281	tcaatgatat	tctagcaaac	caaattcagc	agcatattaa	aaggcttata	tacaaatcacc
95341	aagtgggatt	tatccgtgga	atgcaagagt	gggtcaacac	acaaaaatca	atccatataa
95401	tacactacat	tcataaaatg	gaggaaaaata	aaccacacag	tcattctggg	tgatgcagag
95461	aaagcattag	acagaaccca	acaccctttg	tgatcaaaaa	atactcaaga	aacaaggact
95521	agaagggaaa	ttcctcagca	tgataaaggg	tattttacgaa	aagccaaaaag	tactcatcat
95581	gggtcaactgt	aagactggaa	gcctgtcttc	accacctgta	ttcaacatag	taccggacat
95641	tctagtccga	gcaattcagc	aagaaaaaaa	taaataaata	aaaggcatcc	acattgcaaa
95701	ggaagaagta	acacttcttg	agggaaaaatg	gggaatgatg	gctcgtgggt	acaggggtttc
95761	ttttgggagc	aatgaaaaatg	atctaaagtt	gactgtgggtg	atgggtgcac	agctgagtga
95821	atatagtaaa	aattatggaa	ttatgccc	gaaacgctga	actacagcat	atgtgaatta
95881	cagctcaata	aacccaccat	aaaggaaaaga	ttaaacaatg	atctactcct	tacctctgtt
95941	ctcagagcca	tttccagccc	agcccaggcc	ccaggcaacc	ctctgtgggt	gacgcagcat
96001	tacctgcttt	ctggctgtcc	agtgttgtgt	agcgcaaggc	cctggggaggc	cactgggttg
96061	gtcatgcggg	aaaccaagaaR	gggagcaagt	agttgggtga	caaaacctgt	acatctgtc
96121	ctaactttcg	aaacctggct	gatgggaggc	caatgtSacc	ctctaacgcc	ctcccatagc
96181	agcctcggt	tgtcttctga	cagtggcact	gtccacccac	aggcatcacg	catcctgccc
96241	gcacattgct	tccgccagtg	gagccccagc	tgaaacacta	gacaggagag	aagtcagag
96301	atcctgctca	ctMtagggaa	ggagaaagac	ctcttcagct	tgctctgcct	cggcaaggc
96361	aaagcctcat	gcatttgctc	actcctccct	cccagggtc	agccggcacc	tgaagctagt
96421	gtagcagaaa	gagcccaatt	ctgcatacag	cgggcctggg	ttcaaatccc	agatcctcca
96481	ctttctaatt	gtgccatctc	tgagccttcc	atcccttacc	tagaaacaga	gttctttcca
96541	cccatgccct	tgtgagacag	cgtaagtga	aatgacatcc	acacgggtact	ggaagcagag
96601	ctgctagaaa	cccagccagg	cagggtttga	acttgttatc	tccaaaaagt	ctcaggaatc
96661	gtattttctgc	ctgatgggtga	cctacttcag	caaggaccca	atgtagaat	tactctgcca
96721	gcaaaaaagc	aggcctcaga	ccccctccat	cacctcccYa	at ttatgaat	taaacacaga
96781	ggccagtgcc	tgaggggtct	gcagatgaga	gtcagaaacc	agcagcccc	cacccacag

96841	agacaggttc	ctgctaacta	gacacaagaa	ttggccttct	ggggtgtctc	tgcgtgttct
96901	gtgtatggtg	acctagcccc	ggcctccag	gagggacact	acgtgcagga	catttttgaa
96961	aacagtcttg	gtgtttcctc	aaaaagtga	acactgtgtt	atcatgtgag	ccagcaattc
97021	toctaKgc	atacctaa	gtaatgaaca	tatatgtaca	cacaaagact	tgtccacaaa
97081	tggccatcga	ggagaagctg	gccacagaga	agatggccca	gagaggaggt	gctgtgtgca
97141	gaagcagctc	tgggaggaaa	gacgtgcaca	ggccaaattc	agattactga	cacttttttt
97201	ttaatgcaga	gattgtctgc	agaattacct	atTTTTgact	tatcttgaaa	aactgaggat
97261	ctggcaacac	tgaccccgaa	tcttcctaac	agcaacaatg	atcagaactc	agaggttttS
97321	tgtctcactg	atTTTTtttt	ctttttcttt	tctttttttt	ttttttttga	gatagggctc
97381	gactttgtca	cccaggtctg	agtgcagtgg	tgcactctca	gctcactgca	acctctacct
97441	cccaggtctc	agcaattctc	ctgcctcagc	ctcccaagta	gctgggttcca	cagggtgcag
97501	ccaccacgtg	tagctaattt	ttagattttt	ttggtagaga	tgggggtttca	ccattttgccc
97561	cgggctggtc	tcgaacccct	gagctcaggt	gatccatccg	cctcagactc	ccaaaatgct
97621	gggattatag	gcactttgct	gcacctggcc	ttgtctcact	gattttctgat	cttattttgt
97681	attatctttt	tccttctatt	ttctttggat	caatgttgcc	tttttctaac	tWctcacact
97741	ggaaatttat	ctctttaatt	tttagacctt	tttcttttct	aatgtaaatt	tccatccacg
97801	ttttgatatt	tagtgctttt	attatcattg	agttgtattt	ttgaagttcc	actattattt
97861	cttctttgac	ccatgaatta	tttaaaagta	ttttttttta	atttccaaat	aatggggatt
97921	tttgcttttg	gggctattaa	tttctctctc	acaggcatga	tggctcaggga	acaagacctg
97981	ctgacactat	tcttttagtt	tctgtgagat	ttgcacttaa	ttgaccagtt	gttacaagtg
98041	tttctatgt	gcttgagaaa	aagaaataac	cttacttgt	cagatgcaaa	attcatcaat
98101	gtcaagctta	ttatttgggt	tatttagaac	ttccttatct	gtgctgggtt	ttttgtctac
98161	ttgatccaac	accaattgga	aaaaaatgg	tggattctct	caaaaatgat	gataaatttg
98221	tcaatttctc	tatttctaac	aaagtccacg	tcaaatatgt	tgaggttatt	ttattaagtg
98281	actctacctt	tagaactgct	ttatcatctt	ggggaaatga	accttttctc	attaaacggg
98341	gacctgtctt	tccccagcag	cgtctcctgt	cttgaggtct	gttgagtctg	agattgtatac
98401	agcaatgtcg	gcttgccctt	ggcggcctga	taggtctggt	tccatacttt	actttccact
98461	ttttcttgcc	tttacacttt	actctgggtg	tagcgtccag	ccagattttg	aatttggggg
98521	ttgtctcggt	tgtcagatat	gacagcctct	gtctttgaat	ttgtgagttt	aatccattta
98581	cattttacctg	gattattgat	aaatctgggc	ttgtttctcc	catcttactc	tatgatctca
98641	atttgccttg	gtttttctgt	ctttttttct	cttttataat	ttgtcttata	aaaaagaca
98701	tcccggccgg	gcScagtggc	tcaggcctat	aatcccagca	ctttgggagg	ccaaggtggg
98761	cggatccctt	gaggccagga	gttcgagacc	agcctgacca	acagggtgac	accccatctc
98821	tacaaaaaat	acaaaaaaat	tcgggggggtg	tgggtggcaca	tgccgtgaat	cccagctgct
98881	agggaggctg	aggcaggaga	atcactggaa	cccagagggc	agaggctgca	gtgagccgag
98941	atcgagccac	tgcactccag	cctgggtaac	aaagtggagc	tccatctcaa	aaaaaataaa
99001	ggcatccctt	tttctcta	ttgtcctgtc	accactgtgc	tggctgcccc	agccaggctg
99061	caaccctggg	gacacacatg	cccctccagc	agtgagttca	gacccttctc	accctttggc
99121	ctcaggtctc	tcacttcctg	cccctctctc	tgagagggtg	cacgccatgt	ccacaccaac
99181	tccaaatctg	atcatcggtc	ccagtcacgt	ccctgcacca	gctgctcatg	ggcctttctc
99241	tctgggggtg	ggctgggagc	tgtggtctct	gctgctgtgt	cctgcactac	tccacatttc
99301	ccagttcaga	ggtgatcgga	gccctctcta	atatacgcca	tcccagggcg	gggcagcagg
99361	gcacaaggcc	acttcagtgt	caacacggat	ctcactgcag	gaactaggag	gggcctccct
99421	taggaagagg	cactctggac	tgttttaacc	acagtcccca	cttctccctc	ctgtctcaca
99481	gtcccagact	tcatacacac	acgctgctgc	ctattctgga	attatgcagt	attattattt
99541	tatattcatg	tcgctgggca	tatgatgctc	ccatgacaag	gggccctgga	gcttttagat
99601	caatccttag	aaggactcca	gccagacccc	aaaggccaca	tattatcaat	tgggtttcta
99661	tgaagtatac	agaataggca	aatccaaaga	cacaggaggt	agattaKtgg	tcactatgag
99721	ctgggaaggg	gaatggggag	tgactgctct	gggtgcaggg	ttcccactgg	ggtgatgaaa
99781	acatttttga	actagatggt	gatgatggct	gcacaacact	gccaaaaacc	actggattac
99841	acactttcat	actcatatgt	attatagtcc	attgattttt	aaaaggtcag	aggaagcaac
99901	tcaaatgctt	gtaggaggct	agaagaggcc	cagatgggtg	aagtcggggg	ggcgtgcaSa
99961	ttaagcccac	aagaatatct	tctaagctac	caggtcacac	acacactcYc	tctcactgtS
100021	cctgaagccc	atagccccct	ctgcctgact	ctcctatcaa	ctgtKgatgg	cgtcaaagat
100081	gcacaSaaat	atatttttggc	tgaagaaac	acaacagggc	acaatgagga	agggtcgcca
100141	agcagccctat	tagcagcatt	tgtcatggtt	catctgggag	tctgcctttt	ccccagcca
100201	gactcttcag	ggacagagac	cagattgttc	atactcagtg	cccagcaggg	cgctggcctg
100261	gggtgcttgc	agacactgac	tggatgaatg	gatggatgga	ttgggtggat	gaatgagtgt
100321	gtagatgaat	ggatgatgaa	tgatggatgg	atggatggat	gatggatgaa	tgaatggata
100381	aataatttga	tggatggatg	ggtagatggg	ggatgggtag	atggatgagt	aagtagatga
100441	atggagatca	atgatggatg	gtagatgga	tggatgaatg	agtgggtaga	cgaatgaata
100501	cacgatggat	gaatgaatgg	ataaatgatt	ggatagatga	atggatggac	ggatgacaga
100561	tggataaatg	gatgaatgat	tggatagatg	gatgaatggg	tagacagaaa	gactctgata
100621	ctggaaagtg	aatccttcac	atccatggca	agtgaaccgg	gtaacaaagt	cagaggcaag

```

100681 accatagcag cgtcatttgt cactgttgcc atttatacac atttttgggg cagaatcggg
100741 catgcttgac cacaggtcat ctctagaggg ttaggattgg gggtcaggaa ctgtgggggtg
100801 ggctcagggc tagggctgcc gatgtggcct ttgtgcctgt gagcagactg catttccatc
100861 agcctcatcc ccattctccc tctctgcctc caccagttag ttcagatcct tcctacaatt
100921 caatcttggg gtccctcacc tttctgtccc atctcccaag acctggctgg ggctgactc
100981 agaattgggt cagcaaagcc ctgaggtaga gccgtctccc aagattatgg cctggctggg
101041 gccgcctca cctccattca aaggacacca tgtggaaaac caatgtgccc ccaagcaggg
101101 gtggcacggc acagacactg atggcacagg ggccacttct tgcctcttcc tgggtgtctc
101161 atcctcgggg atcaggggct ctgcagcagc aggcagggct gggcagccct ggagtgtgtg
101221 acacaattcg ggcactgtga ccctactgga cccctttgca cagatgctgS tgggcMtagg
101281 aactggctgc tcccactgg ctgctctggt aggccttggg gtctgcctac accccaaca
101341 catgagcatg ggcaggtacc caccagtctc YttctStgtc cccagcctct cccgagcctt
101401 ggcttgggtc caccgtggac agcgtctctg acacgggtcYt tcttcccaca gggagaccgg
101461 gaggtgtctg ttctgccctc agagcagagg ccccaggctc cttcagggtg gagcacaggt
101521 gcctttcgca cagggcccag cagccagctc agcccacaga acagtaagcc gccagggtccc
101581 aagttgtcat cttgtgacct gccctgggg acacgggtgc ccaacaggaa agaggcggg
101641 ggcttattat aaggaaaata gtgtcaggcc tgccatggag ctttgggccc gaggcctcag
101701 ctgaggccaa ggcagcagga ctcatccaga gacaggacag accctgggct cctgggtgcc
101761 agctcagggg agctcctggg tgcggctcgg tctactccag catttccctac atggatgaaR
101821 atgtggatgt gtctgtctctg accatgggtg tgtgcatgcc tgcgtgtgtg tgagagagag
101881 agaggagaga gagacggaga gagtttatgt acgtgtaact accaaccaga gtccaggagt
101941 ctcacactcc actgcaaacc aggttcggcg ggcagcatg agtggggggg ccaggcacga
102001 atcggcaggg agggaggggt tgtgttcgcc catccagagg aggcacagag cctcccaccc
102061 agtgctgctg gcaggaacac gcccagggtg gccaggcctt ctgaaggaaac tgaaaattga
102121 gattttcatg tgaacctcc tgatttttaa gtgtacactc aacatttctg aagccgttgt
102181 acgagtcaaa taaaacattg cgggtgtggg cggcgaggcc gcccaagagc tgcaggctg
102241 aggcctggct ttagtccagc acccgttgt acatacagac gagaagactg aggtcaggcc
102301 agagcaagga ctccgcaggc caggcccaca ctggcgccc cggtgcctg gcccatcact
102361 ggggtcacca cagcccatgg ggagcaccaa gSacaccaag aaaggaggc ccaggcctgg
102421 tccccggggt ctctgtcacc aagggccca attgtgagtt tcaaagggtg ccgaggtggg
102481 tccacaggc ctgtgagtc ccacactcct tccagttca aggcctcggg acaaatccaa
102541 gccagcttt gtctcgtgg tgatgaaggc caagaatgct gggagctccc gggatattgc
102601 cYaKaacaaa ggagcacagg cacctgacgt tctggacaca agggcagaac cctcttgaga
102661 ccattcatcg ccgtttgctc tccaggctgc gggggcctgg aggcacatcca ggcacgggta
102721 taggtgccgg gagcccgggg aacagaacct cagacccgga actgggcacc agatgcacca
102781 gccgggggtc tgtccggcca gcccggggag tccaggggct gctccgaacc cggcgccacg
102841 cctgtccctc cagtgcgcga tccaacagct ctgtctgcct gggtggggc cctgcRcct
102901 ccccgctgct ggctcccgt gatgcgtac agatgggctt catcgctgcc tccaagaaca
102961 acaggacgac cgcaggtttc gcctcacaga aacggaaccg ctggaaaggc ctctcttgac
103021 attcccgaaa ggaataaaa aattcctctg aagtacattt caactccgag gcagccctgg
103081 ccacatccaa gaagctcggg cctgagcgcc gtgggctcca gtgaaggctc ccagcactga
103141 ggccgcctcc tgacccccat cccggccact gcccggggt gctcaaccct tacacggctt
103201 tcacaaatth tgaattagct gccaacagtt taaaatgaag agattgtaag taaaatcaat
103261 ttccagcttc tcttataaaa ttggaggccc tctgtctgct ggccgacccc gtgtctacct
103321 gggcccagtg atctctgctg cgtcacctg ccaggccccc atcagtgtgc ccatacaga
103381 ccactgtcct cagcgaacgc ctgaccccca ggctccaggc aaggttgaga ggacagatgc
103441 tctcatcctg ggggtctggt ccttccctggg gccagtgtta tccagctctg ccctccagag
103501 ccctctgcac acgctggggc tgggactgcc catggctgtg ggctctgatt tgggtgggca
103561 gacctgagca ggccatcatg cttcttattc ccagcctgac acccccagac cgcccccg
103621 caaccagac aatgtgtact gaactcctcg tcaacttgac cctgagctc ctccccgcag
103681 cccaccagggt gctcctgctg tggtttgggc agcacctagt ctttttcgcc ctgtggagc
103741 tcgctgggcc agacttgca agccatctca gccctttgtt ctctgactc catccctcca
103801 agcctggcca cccagcctc ggccccagct cctgccacgt gaattgtcaa aggcagagca
103861 tgaggaggct gaaggcagcc cccatggcac accccacgga gtccctgag tctccaaaa
103921 aaggctcaag cctgggcagc attgtctgca gctaggaatc tgcagctgga ctcaaatatg
103981 tgtccagaat gatctgggaa aggtccttga gaaaaggggc agacacaggg accttcccag
104041 gcccgatcc cacagaggaa atgattctag cgagcgccac ctgcacaggt ctgaggtcca
104101 gacaggggtg gagctgtcac aggcagccgg gtggggctgt gatctgggtc ctcccagcct
104161 cagggtgatt tggaatggga atacggttct gccaggcaag gagtatcccc ctctgaccaa
104221 ccttggact gcctaccagg actcccaaca ccttctctgc acgatggcg tgaggggctg
104281 atccggccgc tctctcatt cctgtcccca cctgatccag gccgcagtct ctccctgcac
104341 acaggacagc agcctcctgc cggctccctg cctcttccgt gctcaccctc ctcgaaacc
104401 agaccatgcc acctgctgca tggcacagcc cctctccac gacggagccc acagaagcca
104461 tcgtgatctg acctccagtc cctcaggcc ctcacccctg ccctcctgcc acaccccaa

```

104521	acaagccccc	tctcaccccc	tcaggtactt	cgcactctgt	gccctggccc	cttttctctc
104581	tggggaactg	caaccacac	tccaggcctc	agcccaacgt	ccagtgtactg	cgggctcact
104641	ctgtggcact	tcccaagtct	ttcattgtgc	aggtattttgt	ggactgacta	gtgcactgtg
104701	ggctgtatga	ggtcagggct	gtgcctgtct	ggtcccagag	cctggcccag	tgcccgggtcc
104761	tgctgtgcgt	gccgggtaaa	gatctgtttg	atgaatcaa	ggagagagga	aggggagcag
104821	ccctgggatt	ccacaggtgc	tttatgtcca	gagacaaccg	cgcagggaac	agtggccaag
104881	gcacgctggg	agcagtgtct	cagcccagcg	gcattcccag	ggcagaggct	gggattcacg
104941	cattgtctgcc	ctcaggcacc	aaacgggctt	gtggcagtea	caggggcaca	ttcctgatac
105001	cgggtctgtc	gttccagcat	ggacagggcc	atcttctctg	ccatctgcag	tctggcatca
105061	gacattccca	aataaaggac	agacagttaa	ctgtcagaaa	gtgctgggtc	ttgggcctgg
105121	aaggaaggga	gcccgtgtct	actgggcacc	cccaattttg	gccactgtgc	agtgcctctg
105181	gtcctcaacc	tcggggcctg	gaccacagct	ccatgagacc	acagggagga	agcagagtca
105241	cagaagaaag	agtgtatga	agcactggcc	acttgatttg	cagggcacag	tgcaaaatgg
105301	aaacgcacag	tcccttgttc	aaaaagtgat	tacgacttct	aaggcaggaa	tgacagagt
105361	ttaaagtggg	cataggggcc	tgccggatgg	gtcacaaagg	ggcgaagtca	gtcctgtctc
105421	agtgtgagga	agtgggacta	gcaaaggccc	agggccaccc	aatgtcccat	gggcccattg
105481	gaccagttag	cacccacagc	tgccaacctg	agtcctgcag	cctgacactg	tggaacaac
105541	aggaggcgct	agtgccaaag	cctccaatgc	ctgcagggag	caactgttgc	tcccgagtg
105601	agctcctggg	cgaccgggct	tcctaggcac	tatgcttatt	ctttcttttc	acggatattt
105661	acgcaccaac	cgtgtgcccag	gcattcctca	aatgctgca	aattccactg	taagcaaaag
105721	attctccccg	aagctaactg	gctaagtgc	ctggagacac	atcctcactc	acctcctctg
105781	gtccccagtt	tctgcagccg	tttaagtggag	ttccaaggcct	tggtgcttgg	aaggccctct
105841	gtccaacaaa	cagatggggg	cacaggctac	cgggttttgc	cagcccaggc	tctgagcggg
105901	gacaggaggg	cactgcactg	ggtgtcagca	gccctggggt	ctaagcctgc	tctccctcc
105961	tgatgcagga	gtcaaaaggc	ctggctccca	ctgtttctac	agcttactat	ctgtgcaacc
106021	gtcggcacct	cagtttctct	atctgtaacg	tgggcacagc	aaaagcacca	accaccctg
106081	ggctgctatta	caagaatcct	aggagatgat	gtacctgcag	cgtgctggca	gacgcgttca
106141	acggtaagt	ctccagggaac	gcgtcacggt	cttcttgcct	cctgctgttc	ccgctactgc
106201	ttggacagt	gtagccctgc	accttagttt	tctctctgct	tccaggacta	cgtgcctat
106261	ttaatgacac	catgaagagt	gcccaccttc	gcgcaggccc	tcggtgaatg	ggcatgtcct
106321	tgaacgcaaa	cttggctcca	acggggcagg	agaaggatca	ttcagactca	gctacagaaa
106381	aacagacttt	tgtgcaacta	acgctaacaa	aactggaaa	gaaagaggaa	ctacggcctg
106441	gggaaaagct	tgtatcgaac	atgacagaaa	gtcaatatct	aaaaagagat	aaagggttta
106501	tcccaatata	taagacccca	ggagataaac	aggcaaagg	cttaggaata	ggagacatgc
106561	agagtaatca	aaaccaggga	agatcactca	tcctcgccag	taaccaaaga	agcgcaaat
106621	agcgcaaaaa	gcaacgccgc	tttgctgccg	gagcagcaaa	atattaaaca	ctaaagcccc
106681	cgtgctgggt	aggatttgg	gaaaccagg	gctgctgtgg	gagccagtct	ggctgtccc
106741	cgagtgggtg	gtctgttaag	cgctgggatt	tcctaattca	tttagtcact	cagttaggtg
106801	cccctgatgg	ctgctcagg	ccagggtctg	agcgcgtctg	ggctggctca	ggcggccatc
106861	tgccgggtca	gcaacagata	cctatttgc	ggaacagaac	ctgaggtcag	ctcccaggat
106921	gtccccacc	ctgagtttca	agagagtttg	caccaggagc	agggagcatt	cagctgtct
106981	cgggcagctt	ttgtgtggg	aacagctgtg	agcgtgtcct	agcgccctt	gggggtgaca
107041	ggaaggctct	gcctgcaaa	gcctagaggc	agccacacac	cctgcccctg	acactcctaa
107101	gtacgtgcag	gccacagttc	ctgcattccat	tcattccggc	cagtctgggg	tccctcccca
107161	cagtggcagg	caaggggaag	ctggaccggg	cctgctagac	aaacaaggat	gcacacacgt
107221	ccatgggacc	ccaaacctca	cactgtccct	gacagcccgc	cccacagagc	acagcctctg
107281	cctcccggag	caagaggag	tggcagtgca	tcccagggcc	tggtgtcctg	tggggtgca
107341	ttgggggctc	ctgccccgc	cccgtgagtc	tggaacaca	tggtcttgg	ccttggcacc
107401	ttctttcctg	ggaaaacacc	tcctctctgg	ccagggcagt	gcagagcagt	ggaagcgggc
107461	agctgtggag	tcagagtgat	gggtgccaat	tctgactcct	gcacgtggcc	cagaccact
107521	aacctgccac	cccaccagc	acacctgcc	ggccacccc	accagcacac	ctgcccggcc
107581	agcctcaggc	caccagccag	cctcccagcc	acgtgacctc	tcaggggcag	gatgtgtcct
107641	tgattccttc	cttcacacca	cgtcccggaa	gctccacctc	aaactcatcc	tgaaccact
107701	gcctccaccg	ccatctgtca	ccgggactcc	tcctgtcttc	cacgccgtcc	accgaaaacc
107761	ctccaaaggc	ctccaccgca	cagggaatga	attctcacc	cgtaccgtga	ccgtcccttc
107821	aaagcgatga	gtgccaccga	tcctcccg	cctgcacgc	ctcccactcc	ccattacgt
107881	cctcagttcc	aggtaactag	tcttcttct	gccccttgg	cacgtgcgg	taactccagc
107941	ctcagggtct	ctgcatgtgc	tggtccctct	gcctggaaca	cccttcccag	ggctgggctt
108001	tcccatcatc	accaacctgg	ggcgccctcc	tcctttacct	gccttctcac	gctgcccctt
108061	caccccgctt	tactttgttc	acaggattct	ccactgtctg	cgagcatctg	ctcaccaggt
108121	ctccccctgc	gggctgtgag	cccagagaac	agggactoga	tcgggttctc	ccgctgtac
108181	agccccagca	cccagtcctg	tcagacacat	ggtggacaca	tagcaagaac	gtgacagagt
108241	aatgcaggaa	aatacacttg	tggcagcagg	caaggtgccc	tctctgagcc	tcagtctccc
108301	cctaagcagt	agagctgcta	atgctattat	tcccccata	tttggggagc	agtattatct

108361	ccccatgatt	tggggagcag	tattattccc	ccatgatttg	gggagcagca	ttattccccc
108421	atgatttggg	gagcagcatt	attccccc	gatttgggga	gcagcattat	tccatgattt
108481	ggggagcagc	attattcccc	catgatttgg	ggagcagcat	tattccccc	tgatttgggg
108541	agcagcatta	ttcccccatg	atttggggag	cagcattatt	cccccatgat	ttggggagca
108601	gcattattcc	cccatgattt	ggggagcagc	attattcccc	catgatttgg	ggagcagcat
108661	tattccccc	tgatttgggg	agcagtatta	ttcccccatg	atttggggag	cagcattatt
108721	ccatgatttg	gggagcagca	ttattccccc	atgatttggg	gagcagcatt	attattgagc
108781	attaatgaac	ccgcatggca	ttcaggtgcc	cggccttttc	ccaaccagac	cagcgcacca
108841	gccacctctc	cacacaggcc	cagtgcaccag	gtcacctcct	gtcctgtcct	gcaggtctct
108901	tctgggtgtg	ccgacagttt	gggatgggag	gcggtgggag	cgtgagctgc	gctcctggcc
108961	gggttctcta	ccagctgtgc	tcgtggcgca	gcccacacct	cacggcaagc	cccagccaaa
109021	tgtagccaagt	caagagctct	gagagacatc	ccagccgaag	aatccagccc	ttgtcccaaa
109081	cgtgtgtatg	cacaggcaca	cgaaagaagg	tgagacagaa	ggagagaaaa	atgaagccag
109141	agagcaggag	agagagaaga	cagacgccag	cacgcccata	tgactgcgcg	acctttcccc
109201	gctcagagct	ccttcagggc	ccagagcaca	gcgctgccta	ggattaattt	caaagagaaa
109261	cgggtcaaac	aaaattggaa	atttttatag	aaaatatttc	aggcttggag	ctggctgaag
109321	tgtctgccaa	cctgcagaga	attctgagag	gagatggggc	ctgtggtcag	aggcctccgg
109381	ctggcacaaa	ctggcagagt	ttggacctcg	gacccaaatg	cagtttctga	gccaggcctg
109441	tcctcgctgc	tgctggggaa	gagcttcctg	tccccttcag	atgtccctac	agctgcctct
109501	gcagccctga	tgctgggtacc	caacgcctct	cacctgtcgc	cccacccctc	tctatcccca
109561	agaggcctct	gtctagggac	ccactatggg	atgcagtccc	tttgtcccca	ttatgcaggg
109621	tgccccatgtc	tctttcacccg	tcattcatgg	ggcagccagg	gcatgaagga	agcctcacct
109681	gggtcccaag	caggtcccac	tgctccttgc	ccaaacttca	ccccagccag	caggctccta
109741	catccaaggc	acagatgtca	cagacatgaa	aagatgccct	ggacacatgc	ctctctccca
109801	gaagcctgcc	tctgccatgc	tggggacctg	aaaagagggg	gcatgccca	cccacagttg
109861	ctgatcaaac	gccagatgcc	cccgtgagca	gatcacagct	tcaccctga	gcccgtgtatt
109921	gtctgcagtc	cagggactgt	tacctgtgtc	aaaacccgaa	aagctctttg	tgggacattc
109981	ccaacacaga	tccaagggaa	ggaaagaaag	tatcaattca	ctgtgtctgt	ggaagatggc
110041	gaggcagagg	agggaaagg	gacggagagc	aagtgtcgca	gtgagcactg	ggcaggcgcc
110101	ctgtgccctg	cgtaattcca	gcagcccctg	tgccactgca	cccagggtaca	gtggagatct
110161	cagcaccacg	gttctgagat	cctgaggggc	tgggtgagct	gctctgggcc	ccttagtgag
110221	ccagaacaac	agcatgtcgg	gtggagctga	gcacactgac	ccctgcacca	acagcagcat
110281	gctgggcgga	gctgagcgcg	ctgaccctcg	caccagcgct	gacccttgca	ccaacagcag
110341	catgccgggc	ggggctgagc	acactgaccc	ctgcaccaac	agcagcatgc	cgggtggagc
110401	tgagcacact	gacccctgca	ccaacagcag	catgccgggt	ggggctgagc	acactgaacc
110461	ctgcaccaaa	ctgcccgtcg	cccacccac	tcaccacact	agtgtggaa	gcatctgtgg
110521	gagcagccct	aatgcagctc	ctcagaacca	ctggccggtc	ttccagaact	gtcccctgca
110581	ctcgagtttg	ggctgcagcc	aagccagggc	caaggccacc	atcctccggg	tcaggcttgt
110641	cacctgcccc	cggtgtctcca	ccaaagatcc	aggcccttgc	cagccacatt	tacctggtcg
110701	gctccacccc	tgaaccagaa	accatgtcgg	gaggggagag	gtggcggggg	aagggcaagc
110761	ctgagactca	gggactaaat	ttatgtcgga	ggcacagggg	agttgtctcg	cacgtggggg
110821	ctgtcaggcg	gagtgccagg	cccagacgcc	gtggaggggg	agcacaccct	ggcacgtctc
110881	cactgtggaa	taaggggggc	ccagccctcg	ctggggctgc	aagcctcgct	tctgccttca
110941	cagcccaggg	ggtgggaagc	acccacaggg	cagcgtctcg	agcagcagac	tccagccaga
111001	gagccagcca	ggcccgggag	ctggcccctc	ctgctctccc	cggctgacct	caggcaactc
111061	cagcctcgga	gaatgtgcag	cctcggtttt	cccagcgccg	gcccccgccc	acccttccg
111121	cgtggggccac	atcacccgcg	tggcatctcc	aacagccatc	ccaaggggcc	ggtctgcacg
111181	aggacaggag	cgccgcggcg	tctgccaagg	gtgcgctcag	ccctgaagca	cgccctccgg
111241	agccgcctcg	ggtttcacaa	tcagtgtctc	tccctgggtt	tcccggggcc	tgggctgcag
111301	gctcagctcc	aggggtggaa	gagaggatga	gctgagagga	ctccagggat	ccagcgccga
111361	ggcgctgaac	tcagttagtc	ggcaacagtg	gcttggcagg	gcccaggctg	accagggtat
111421	gagagagagc	acctatcacc	acgccagagg	ccctggcctc	aggggtctgg	ggagtcaatt
111481	gggaagtgcc	cggttggaga	cacagagaca	ggtctgggca	caagacgctg	gccagggtcca
111541	gggtctaccc	tgcaaatcca	gccaaagccc	tctctccatt	ctcccttagg	acccaagggt
111601	caacacttag	acatgtcagc	tgaatcatga	ccccggagc	tccatgtcct	aatcccccaa
111661	aactcgtgag	cattacctta	tatggcagaa	gagactttgc	aaatgtgttt	aaatcagggg
111721	tcttgagatg	gggaaagtaa	cctggatggg	cccagcgggc	cctaaatgtt	atcacaagtg
111781	tccttctaag	agggagagct	ggctatagaa	gaggagaagc	cacgtgacca	gggaggcaga
111841	gactgcagca	atgttgccac	aagccaagga	acgccaacag	ccccaggag	ctgggagggg
111901	ccaggaattg	attctgccct	ggggcctccc	gaaggaaacca	gcctgccag	cacctgtatc
111961	ttagccccta	aaatcctatt	tcagagttct	gagctcccga	actgtgagga	atgcaggtgt
112021	gctgttttaa	gccattaagg	tgtgggtaac	ctggtacagc	agccacagga	agctgagatg
112081	gacgtcgaca	gctacagggc	agaaaaggctc	tctgacctgg	cattcagcaa	attggcacc
112141	cccaacaaag	cacacggccc	tcctgcccc	tcctggggccc	cccagccccc	acctgggcca

112201	agcctcacat	ccgctgtgtc	agtggggctg	gctatctgct	cgccactgaa	tggatgagtc
112261	tcgcctgctc	cccagtcctg	tggctctgaca	gaacagagct	cctctctctc	tcctgtgcca
112321	ttgctccagc	ctccccacc	agcctccctc	acactgcaac	cagtacgagc	ttgctgaaac
112381	actcaactct	ttgtgagact	ggaggggcgc	atccccacct	tgggtcattc	aaggctgctg
112441	cttccatgga	ggctgcccc	cagctctctg	cctccagccc	tgggctctgc	acaccaggg
112501	ggaactgtca	tgcttgaggt	ccgccccac	ccccgcgc	cactcttccc	ccagatccc
112561	ctgcggcctg	gaaaactcct	acccaactca	aggcaaatgc	acagggtgga	gactgtgggt
112621	caggaggggc	tcttcttgcc	ctggtgggga	agggtggtg	gtgggctgca	ggtggcagcc
112681	acgctcagcc	acatggagga	gctgcccagt	gccggcgctt	gtcctctgtc	tcctggaggg
112741	ccttccttgg	tgacgcgag	ctgagaggt	cccgtggtg	cgtgcctcgt	ctgcccgtca
112801	gggaggatgc	tgacgcctga	cagcagcag	cagaggcttg	gcaccggcgc	tgtctggggt
112861	ggaggcagcc	aaggctcactg	gcacgttgct	gtctgacttc	tcttcttcca	gcccagcagg
112921	gtgtgtgcac	agcagcctgg	tagcagtcc	ccaggatggg	aagcacctgc	ccctcctctg
112981	tcccaaggcc	aggccggggc	gagccccagg	agagcagccc	acagaggagg	gcagaagatg
113041	cttctacctc	tgcatcacag	aaagcatggg	tagccaatgg	gggaccgtgt	tttgtttggt
113101	ctatatagtg	agcacgggag	gggaccccc	gcacgacagc	tcccactggg	tgtgcagggc
113161	ccagggtctg	aggcagagag	ctgatgtgga	ggcttctgtg	gaaagagggc	tttgtctgga
113221	tttgcctggc	gagccctgga	cctggccagc	gtccttctgt	caggacctgt	ctctgtgtct
113281	ccaaccaggc	acttctctgag	tgactcccc	gacccctgag	gcctggtcct	ttgcgaggt
113341	gataggggct	gtctctctga	cccgggtcgg	cctgggccc	gccaagccag	ctgttgccga
113401	ctcactgagc	tcagcgcctg	ggctggattc	ctggagcctt	ctcagttcat	cctctcattc
113461	cagaaatttt	aagccacatt	ttaacatcag	gaaagtccac	ataaaaaagt	tgatgtctgg
113521	attttcttga	taggatgttc	agctgcacca	gacccaactt	ctaccaaatt	gaggtcaagg
113581	gagttgagga	gggtctgtcc	ctccttgca	ggcctcccag	acgccctgtg	cccatggcaa
113641	ggccagggct	ggacaccacg	tgtggctgca	ttcccactgc	tgtttctttt	tagtggagat
113701	gtattcttgg	tctccacgtt	tcacatcaaaa	ggagagaaat	gaaagccagg	ccaaggggtg
113761	gtgcaattat	ctttttttcc	tcactgtctg	tccttgttgg	cattgcctgc	caggcctgc
113821	aggctctgag	tttgcagctc	ttgtctttct	gcaaatggca	ctgtgtgtct	gaaacagaga
113881	ccacatgctc	aacacgtgac	tcggggtcct	tctcactgtg	tggcaacctt	cacagagtcc
113941	caaggctcag	ctctgggaag	ctaagacag	ccctcagtgt	cagcacaggg	ttggcttctc
114001	tgtgcacgtg	gtcttgaggt	cagcgcctcc	aggtccctta	acgcgggcgc	tgggggagga
114061	ctgccaggct	gagactggca	cttagacatg	cagggaagag	ctcccacacg	acagggtgga
114121	gggtccagtg	agggagggcag	caggaccagg	gcctgcctgg	gctgtgcctc	tgggctatat
114181	gcctggggga	ggcgtaagtg	aagcctacag	tgctgctcct	ccgcgtgacc	cagacgggcg
114241	cctcccgggt	gtgctgtgca	ctgcgtggcc	agccttgtag	gtgagggcca	ttcacatcct
114301	gctgctggct	gcgtgggtg	cccaaagatg	ctggatcctg	atccctgcaa	cctgagacgg
114361	gtaccttcta	tgggaagaaa	gggctttgca	gatgtgcttg	cagtaaggct	cttgagatga
114421	gaagcctgtc	ctggattact	cgggtgagtc	ctaaatgtaa	tcaaggcatc	cttgcaagag
114481	ggaagcacag	ggacacgtca	gaagaggagg	ggcgacatg	atccgggggc	agagattggc
114541	atgatgcagc	cacaagccaa	gggaggctgg	cagccccag	gagcggggat	gggccagaaa
114601	cagattcttc	cctggagcct	ctggaggag	cacagccctg	tccacccttg	atcttggctc
114661	agtgagattc	atttcagact	tctggcccac	agagctgtga	gagaggcatt	tgtattagtc
114721	cgttttcaca	atgctgataa	agacataccc	gagactgggc	aacttacaaa	agaaagaggt
114781	ttattggact	tacagttcca	cgtggctggg	gaggcctcac	atcatggcag	aagatgaaag
114841	gcacgtctca	catggcagca	gccgagagag	agagagcttg	tgcagggcaa	ctcccgtttt
114901	taaaaccctc	agatctcgtg	agacccactc	accatcatga	gaacagcatg	ggaaagacc
114961	accccatga	ttcaatcatc	tctcactggg	tccctcccat	aacacatggg	aattatggga
115021	gctacaagat	gagattcagg	tggagacaca	gagacaaacc	atatcaatgt	tcctattgtt
115081	tcaagctccg	tgaggactcc	ctcctcaaca	gtcccacaca	ggatgagatg	cacacacacc
115141	agagaggaca	atgccacagg	gtgccctgtc	cctgtacccc	agcacagtgt	ccctggacac
115201	agagctggcc	ttgggaagg	cagagaacag	ggattggaga	tgctgtcag	aatgcctggg
115261	gggacacctc	aaccactgct	cccaaaaagg	ccaccagacc	tcagagctcg	gcctccggtc
115321	ctggccaggt	tcccaggcca	gcagcactga	gcaggaccct	tctgccaagg	ccagcatggc
115381	caggaggagg	agaaaaggac	cctccccgct	ccctcacccc	ttgcccctcc	taataacatc
115441	ctctcccggg	tgttgccctaa	gctgcccagt	gcacacactc	atccctggcc	tgtctgtgaa
115501	atgctttaaa	accgtcattt	tctatccgtt	tccccaggc	aatctgcaat	ccagcaactg
115561	atcttggatg	acagagacta	ttagtgaata	tgattccctg	aggctccatg	agtccccgag
115621	atcaccctga	gcagcccgtg	ggcttctcgg	ggaaaacagg	acatccctgaa	aaggaggcac
115681	ccctgacgct	gcagggatat	cctgccaccc	agacagatgc	ccacccccca	ctgggacacc
115741	cagggcaggt	gtgtcacctt	ccccggcaga	acataaatag	cagggtgctt	cacaccact
115801	tccagacgtt	gcttttgggg	aggcaggaag	aaactgccag	ccagcacctc	ctgcccagctt
115861	ctgccagggc	cacactgata	tatcagggtt	ccctcagaat	tcactggaag	aaagggtccc
115921	ctatacacac	acacacatac	acacacaccc	gcaccaggt	cctggagttg	tgtgccaggc
115981	aggctgtgcc	cctcactatg	cctggccaca	ggacacagct	tttatgtatc	tcccagttgc

116041	cactctgggc	cctgaaaagg	cccaggcaga	tggaaaagctg	gacgcctgag	gtcaccgtgg
116101	gccctccacc	cagctccggc	actcactggg	cctgggtgcc	ttccagtcag	tgtgtcgatc
116161	tcagcccagg	gccacgtggg	aacgattagg	tgcattggcg	gggagctagc	tcactgaagg
116221	cccagagccc	tcatgcaggg	tctgctgaca	ccagacagga	acatcaagca	aaggggactt
116281	ggccccagcc	ccaagttctg	ggatccagac	ctatagcccc	aggccccacg	tcacatccac
116341	agcctgggccc	ccaggcctgg	agctaaggct	ggacaggctc	acctcctccc	cgacagccca
116401	ttaccaaacc	cccaccttcc	agcccagcct	cacccctgac	ctggcagctg	gcctcaccca
116461	gcgtcttctc	aggccaatgg	tttctgggct	tattctgaga	ttgaaattcc	tttctgtgtg
116521	tctgaatttc	aagttctcat	taaagatatg	tggattgggg	gtgggggtgg	agggaggatt
116581	cagcagaaga	tgccaaacac	cagcagaagc	ccagaaaata	aggcaaggcc	ggccaggctg
116641	tgggggaagg	gctgccatgt	tcaggaaagc	aagcaccata	tggaatctgg	gatcagagac
116701	atggccaagg	gcacgtggcc	agctcacctc	acctgagcag	aggcccccag	aaccaccctg
116761	aatcttccag	gagaggccag	ccagaagcac	caagcccaca	gcacaagtga	ccagcccctg
116821	agctctcggg	gacacagggg	ttgctaattg	cactgggctt	gcccgcagac	agctagctag
116881	caaggagcaa	agctggggcc	ctgcccccg	gtctggagca	ctcgctcagc	cccttccacg
116941	ctggcccttc	tagattactc	ctccttcctt	ccctgcaaaa	tagtccaagg	gtggggataa
117001	gtcgaggctg	agaccacagc	tggaaaaggc	tgggcaggca	gtacccagtg	gcccggctgt
117061	ctgcagcata	gcagacgctg	atttgtcacc	aaggaacaag	gggcccaccg	ttgcaactgat
117121	tggctccaga	gtacattggg	cttctgatg	gaacaattcc	tttgcaagg	tatctgatgc
117181	cgtttcatga	aattagaagt	atgcctgcc	ggctgtaccc	aatttttctg	tcaagctagg
117241	aagtacagca	gacctacagt	aaatgaggcg	gggctgctat	ggagtgcagg	gagtagctta
117301	gtaaatcaca	gcacgggggc	tggcccagcc	caggccccctg	gagcactcgc	tgtgaacctg
117361	cagggtgtgcc	cgatcctggg	ccccagactc	cagactctcc	tcacccaggg	gtcctgacca
117421	gaggctgagg	gctgagctcg	cagcttccca	ctctcccagc	aagagccaag	cacaggccag
117481	gagtccttga	ccacacgggg	gacctgaga	actatgggccc	tgcaagaacca	catcagacac
117541	caaccctggc	ttcaccacaa	tcaggacca	gagactgggg	gatgctggga	gggtgtgggt
117601	ggcacagtga	ggccagatct	gaggcgacag	tcattggaactg	tcaccatggc	cgctggcgag
117661	ctcctttcac	gtttctgggc	ccagggcaca	cgtgggttct	gagttgaaaa	agacagacgc
117721	ccatgagtct	gtgcagagac	tgctgaggct	gctgggagca	cctccagagt	ccttgtaggg
117781	cacatgcctc	cagggatcca	gaaagagcat	gcagggcctt	tctccctgag	cagccagggtg
117841	aggggttggg	cctaccacag	gcacctccc	agggaccatg	gcagcacaga	gcaggagagg
117901	cgctgtctgg	ggctctggag	gcctcttcac	aggcacagac	tcagagacgc	aggcactgga
117961	ctcaggcctg	gccggggggc	cccggcccg	tcacgcacac	ttgcttgcc	ttaggcaaca
118021	gaagatgggc	ccttggacag	ccttcatagt	ccaccatgga	ggccccagac	atgaaggcag
118081	aggtgtctgg	gctgcttgtc	acccacatgc	accagaaca	gaggccccag	ccccagcagc
118141	cctgctacgc	cctgcggtca	gccagcgagc	tggcccat	ctgagccact	gagaccgggg
118201	ctcagcctca	gcagccaagc	tggctctgtg	agagctgcct	gagtcctctg	gatgctcccc
118261	tggggaccag	ggcctcagag	ggctctccac	acagcctgcc	cacctgcagt	agcctcacct
118321	gtgtccaggg	cctgtggccc	cccagagagga	ggggacgtgg	gtggccggcg	atacaccaca
118381	tgacacggc	cttgcctcagc	ctcctgcgcc	gccagcccct	tctccaaggg	ttcgatgaag
118441	aactcctcct	cctccatccg	gatcagacca	gcctgcggga	caaagacaac	aggatcagat
118501	ttccagcaca	caaaagacca	aggaagggga	acagcgagga	gacccccctca	gggagtacct
118561	cccacctgaa	gacctgccc	agcaccaccag	ctgggtgctca	gcttgaaccc	tgggctggcc
118621	ggagggtatg	aacatgttgt	cccagaggcc	gccccccgccc	gcccaccctg	tctaggaaaag
118681	gggcctctgt	ctgggtcatc	tccaccccta	tacaccaacc	ctgtctgctt	tcagaaaagt
118741	tactgtgttt	ataaagtcaa	aaccgtccag	agaagcaggg	agacattcaa	gggcagctgt
118801	gggcctgtgc	gtgtccccag	gagggcatcc	accaaagtgc	atcaatatgg	gccaagtgtc
118861	gccccactc	ccaggagccc	agagtccaag	acaaaggcca	aaacagccag	ccggggattg
118921	aatctgccca	cacacatgtt	ccatctgggc	caaacaatat	ttttaaagct	ctgaagctaa
118981	cactcagaat	gatgagattt	catgtcttga	acatggaatc	ccagctttcc	ttggaaacgc
119041	aggccctgac	cacagcgggt	cttggatgga	acacatgccc	ctgcaactgg	ctccagctcc
119101	tactgcccc	caactgtctc	tcctgtctct	tgtgtcagcc	accagcccag	actctacagg
119161	cagctgaact	gctgcccctg	accccctagg	ccagccagca	ccctccctga	ctacaaactc
119221	ctcgctgggt	cctcatattc	gttgggactc	aggctgtct	atatgctggc	attcaggctc
119281	ttggcctgcc	ccacccctcc	agaccgcggc	ttgcaccacc	ccatgcttct	cactgggctg
119341	gggtgaccct	gccgagccac	ccagccctg	aaccccatgc	tgccaactca	ggaatcccc
119401	ttctcccacc	ccaccacccc	atccccagct	cggccatctg	gtaaacctgt	gttccccctt
119461	cgaggcctga	ttcagacccc	ctcaactctc	ctgcccctcc	ctgccatcca	gcctgcccctc
119521	ccctgccata	ccctgccttc	ccctgccata	ccctgccatc	cagcctgctg	gcaccatcag
119581	gtccacccat	cccattgaaa	ctccagctcc	tacactttta	tcttctggtc	cattatccat
119641	aaaccaaggt	ctcacactgg	aagtgccacac	tgtgtccagc	cttcagagtg	tttctctcag
119701	atcacaagct	tatgacccat	gcaggggatt	ttaaaagctt	gaatttgttg	ccaaacttaa
119761	gaaacaggag	agttcacata	aacgaccaga	tttccagact	ctcttgaaaa	tcagaaaagc
119821	tggcatttcc	agcaccatc	gacactccag	cttagctggc	gctggccggg	gtcccctgtc



119881	cagtttaggag	tgtgccctct	ggtgggcagc	ccccggcacc	tgccatctgg	atgtgcgga
119941	accacacccc	ccccaaagta	acctcatgat	ggcctctgcc	atatcatcat	cttggttqtg
120001	ggggggcacg	gtggcgacag	gggtgcccct	ccttctccct	gaccccaggt	gctcggtag
120061	ccccatcctg	ttgctggctg	ggagcctctg	cagaccgaga	tgcagtcagt	gccccccagg
120121	gcctgctggg	cggtctgccc	tagcctcaag	tgtagcgagg	gaggctggag	atgcctcccq
120181	gcgcactcag	cctgggccc	cggaagctg	atgtgccagg	caggggctgg	tggaggtgca
120241	gggcctggag	gaacacagag	cccagcatca	ggcctggctc	cagaatgaag	gatgcctccq
120301	gcaggaggtg	acatctaccc	aaaccacccc	aatccctttg	tgctggccag	gcaaagagtc
120361	agagagagga	ccagtgtccc	atatggggga	aaccgcacat	gccatggccg	gggaggtgga
120421	cgagacacag	ctggtgggaa	gagctgacca	ggtgggtgga	caggagcatc	aagggcacgg
120481	tgcccggtgg	cccatcactc	agggggtctg	tgaggagat	tgtaggcagc	ccagggcttc
120541	tggggggagc	cttgatctga	ttccagttca	gaagcatggt	ggctgcggca	gggggaatqg
120601	gctgcggaac	aaaaatgggt	ctggggtcct	ggggagaagg	tgttacaggc	acctgggaga
120661	ggggtaatqg	tccagacggg	agcatgagag	acaaggacga	gtggcaggt	tgaaggaggt
120721	gaggagcgaa	atggacagga	ctggttaggg	tagagccggc	tgagatggga	gggagggagg
120781	agcacctggt	gggcttccag	gtctgtggct	gaagaaccgg	ccccaaaagt	taccaggcc
120841	cccatccctg	gcagctgtga	gggtagtggg	catggtggaa	agagactctg	cggtgtgtat
120901	caggtcagct	ctggggctgg	ctgggccctg	aatgcagtcc	ttggccactt	ccttatgccg
120961	gggagtgagg	ggggattagg	cagatgcgca	gaggacgagg	ccatgggcag	acagggaggga
121021	tgccagatgc	cagcttttga	gattggagcg	atgtggcctc	aagccaagga	atgctggcaq
121081	ccccaggag	ctggcagagg	ccagaaagtg	tcctcccca	gagtctccag	agggagcgca
121141	gccctgccga	tgcctggatt	tgggtcaggt	tataccgact	tcggacgtcc	ggcctcagaa
121201	ctgtgagaga	acgaattcct	gctgttttaa	gtggttcctt	acagcagccg	cagaaaagta
121261	attcagggtc	ctggtgggtg	gccggcaaa	tgggtgccat	tctccgaggg	gcatgagaag
121321	aagggcaagg	ggcgacggac	cctctgggct	caggcatcct	gggcttcttg	ggaagaggcc
121381	taggaggtcg	tggtaaccag	agttaccagg	gcaagaccga	ggtcggaqaa	ggaggtcagc
121441	agggctgtgg	gaggagagcg	gaagaggggg	tgggaaccct	gcaggcccca	cgcttctgac
121501	tcatccctgg	gtctcagggc	tcaaccaggc	ctggcgaaga	ctcagcagat	gcctgtggaa
121561	tgagcgaacc	gatgacccca	aagctgcagg	ccacgtcccc	caagggaggc	accgagcagg
121621	caggcgggaa	gaacacagcc	aggcctgtcc	actggcacag	gggcagtgtg	tttaccagt
121681	caggagttag	gagggctccag	cactgatggt	cccgtgctg	ggtgtggaga	aggtgggtgc
121741	ctcaaattcca	tcttatgact	gaaactcatt	ggatgaagt	caggaattaa	ccctgggggg
121801	cggcagcggg	caagctgtcc	tggagcgtga	agttctgact	tcctggcacc	ctttctatct
121861	ggccaggtga	cctcaggcaa	ggtgtcaagc	tccccaaagg	tccagtga	cgtctgtaca
121921	tgaggccggt	catttcccgt	gcacactcac	acaggagcac	atgcagcctg	ggagacactg
121981	agaaccaggc	cacgcctcca	cccaagtgat	ttcatccagc	cagcttgctt	tccaaatgag
122041	gaaaccgagg	cctagatggg	cagacaggct	gctggggaca	cacaggttct	agggatcaga
122101	actcacacct	acacggaaag	ggcatctccc	agacaggctc	ctgggcca	ctgggcccga
122161	tgagccaatt	agggcaggac	actggggcat	gaggaccctt	tatggcagct	ccatcttctg
122221	ggggaacttg	cagcccagg	cgggagccat	cgatgccaca	ggcctgaggc	cctgggggtta
122281	ttctggaggg	acctagtgtg	acaatgaggc	caggctcagc	ctggggaaag	gcagctactt
122341	ctggccggag	taaccggggg	gacagtgttg	ggtctccctg	accagcaggg	gtcaggctcc
122401	cctccctggc	tgccctccctg	aggacaatca	ctgggctgag	ccaacagcac	tagctccatc
122461	tgatgctgcc	ctgggccttg	cctcctggaa	gggcacccct	gcaggcgact	cctcggaggg
122521	cacaggagca	cactggggtc	catctgaagc	agggcctggg	ctcaggcttc	ctcacacagc
122581	cagaagcctc	caagcggtgg	cattctccaa	cacagacgtt	caacatgctc	ctcacagccc
122641	cqcccgtctc	tgttgaccac	cgtgcgatgc	gcacaacgca	tgctggacga	tgaacatggt
122701	ggcacctgcc	ccgagaccaa	aggctgaccc	ccccgagacc	aaaggctgac	ccccccccga
122761	gaccaaaggc	tgaccccccc	cgagaccaa	ggctgaccgc	ccccgagacc	aaaggctgac
122821	cccccccccc	gagaccaaag	gctgaccccc	caccgcagac	caaaggctga	cacccccca
122881	agaccaaagg	ctgaccccc	cgagaccaa	ggctgacccc	cccgcgagac	caaaggctga
122941	ccccccaccc	gagaccaaag	gctgacaccc	cccagagacc	aaaggctgac	ccccccgaga
123001	ccaaaggctg	acacccccca	gagaccaaag	gctgacaccc	ccgagaccaa	aggctgaccc
123061	ccccgcgaga	ccaaaggctg	aacccccccg	agaccaaagg	ctgacacccc	ccgagaccaa
123121	aggctgaccg	cccccgagac	caaaggctga	cccccccgcg	agaccaaagg	ctgaccccc
123181	ccgagaccaa	aggctgaccc	cccgagac	caaaggctga	ccccccccga	gaccaaaggc
123241	tgaccccccc	ccccgagacc	aaaggctgac	ccccaccccg	agaccaaagg	ctgacacccc
123301	ccagagacca	aaggctgacc	cccccgagac	caaaggctga	cccccccgag	accaaaggct
123361	gaccgcccc	gagacaaaag	gctgaccacc	cccagagacca	aaggctgacc	ccccgcgaga
123421	ccaaaggctg	acccccccga	gaccaaaggc	tgaccccccc	caagaccatc	agctgacgtc
123481	ctgccaagac	cgctggccag	aataatgctc	ggggcggggg	gcacttcacg	gaagaagtgg
123541	agatcatttt	attccctgat	tttcaaacac	ttcagtatat	gcatcccagg	aatgatctgt
123601	ctaaaacaca	cagaagaaaa	ttgacaggag	agcaaagcac	aaaatgtgag	ctatgttttg
123661	ctgcattttt	ttccgggact	ttgattttcc	tgctgcta	gtcaagtgt	tctgttttta



123721	catgtaattt	ttattaatta	gaatcagttt	cctctcccca	ctcaataaat	attaaagaaa
123781	ctccagaagc	acactaggct	gaaaagcact	cagttgggag	cctgctggtg	gcctgtctgt
123841	ccaacccac	agctgtccag	cctggggtct	tacagaaagg	ttggtacccc	gacctggctc
123901	taagtgtgag	tcccgctctg	ctcttgaagt	ctgtggatgt	ctggcaaac	cagccagggt
123961	ctcccctggg	ggacagagag	aggatcatgg	ggagggtct	gtccctgacc	ttcagggtgcc
124021	accgcagggg	cccagagcct	ggtagtaccc	gggtcatcaa	gtcaagttgc	cggcattcag
124081	atcctggtga	gttacacaac	gcttgccctg	cctcttcctt	ctatagaacc	tgctcctttg
124141	cagttgggatt	gatggtaaag	tagttactag	aaatggcaac	tagatgttaa	aaattttataa
124201	agatccaaag	gcttaacagc	tctcctctat	tactctgaat	ttttaagtat	tcaaccacag
124261	acagcctgga	agtttaggggt	tcacaggaaa	cacaagcatg	catcctgttt	ccttctgcag
124321	agcaccagg	ctggctgggt	cccactgccc	cttcccatgg	aagccacagc	tgcagccatc
124381	agcagccccc	cacctgtgta	tgactgggac	ctggtctgtc	actcaaccct	gtcaaccctc
124441	agcataaaca	ctgtcagggtg	tgtgccaaagg	ttgcagtggc	aacaatcaag	catagaaggc
124501	acaaagcata	ggaagcacac	aggtccatgc	acagagggac	ccagactttg	ccaagggaca
124561	aggctgcaag	tccctgtttgc	cctcaaaactc	acctccctcc	gaaatacaca	cacagcacgc
124621	accttgccaa	gagtcaccca	gatgctggca	tctgctaaga	ttcccatccc	tgaaggtatt
124681	ctaaactccaa	gattctttct	cccactcttc	cttttggtaa	tgaacctggg	agatctgctg
124741	gacaacgggg	caatggccag	gcagcctccc	ctgctgtgac	agctgcctgg	tggaatgacg
124801	gtctcacagg	gagagctgca	cagtgcctgc	agttcccagg	gcgccctctc	tcagagcctg
124861	ctcgggcaact	ctcttgccctc	tcccagctcc	ccaggagggt	ctctaacaac	cttgggagct
124921	ttcagggaag	ggggctggac	ccctcactac	tgaacagtcc	caggagtggg	aagtgtcatt
124981	gagtgggatt	tgagccactc	cgtctccacc	actggacagt	cccaggagtg	ggaagtatcg
125041	agtgggattt	gagccactcg	gtcctcacca	ctgggggctc	ctttacagtt	cagcatccca
125101	gaggatttca	cacactgact	tttccacttt	tccaaaaggg	acgccagcgg	gcaggagctg
125161	tccatgggggt	taggaaacag	ctgtccctcac	ccgcactctc	accagctcct	cctctgtctc
125221	atggggtcct	cgccagggtgc	cagccactcc	cgtccttctc	cttgagagc	gaccccttgt
125281	ctagggtctt	gtcttgagc	tatgctggga	gctgggcctt	ccagggttct	gctgagcaca
125341	gtgaccacaa	cagagttgcc	tctgagctcc	acagaacctg	atgactttgc	ttggccactc
125401	taaccttccc	tggacctggt	ttcaggccct	acagaccttt	actccagaaa	gtctcttttt
125461	tggcctaagc	tggccgggagc	caattttccaa	caggaccaag	gagccccgac	agtcacatgc
125521	ccagcgaggc	gctaaggaaa	gctagactcc	actttgaagg	gcagggtggg	ggcgaggca
125581	ggcagggaga	gtaagcactg	gacgtgagc	cagggggcct	ggtttccaat	gcggctatgg
125641	ccagtccctgc	ccccttgca	gccccagtt	ccctcatctg	caaactgagg	agctggacta
125701	cgtaacctat	gaattactgc	ttatcctggt	tattcatcac	ccccaaaaaga	aaccccgcac
125761	acgttggcag	gtaccgatca	tccttctaaa	actctactag	gaccaaggag	aaagaaccaa
125821	accaaaaacc	agaaacccac	tctgtataaa	taaaggccaa	ggccgtccca	tggcgagcaa
125881	ggccttccat	gggggtccag	cagccctctg	tccttgggag	cacccccagt	gtcttctct
125941	ctctccgggc	ttctatattgc	acctcaaagc	ctacaccact	gcccttctct	cttggtatag
126001	aggtcagtaa	gtagcagaga	gctggtaaa	atctgtcagc	accaaactga	gactccccct
126061	gactcaacca	gagagtttaa	aatgcaactg	aaaagggaat	catttttcat	tataaagttc
126121	aatgcttttt	ttcttctttt	tttgagacag	agtctcgctc	tgtcaccag	gctggagtcg
126181	agtggcacia	cctcagctca	ctggcaactc	tgcctcccat	ttttaagcga	ttcttctgcc
126241	tcagcctccc	aagtaactgg	gattactggt	acacaccacc	acatccagct	aattttttgta
126301	tttctagtag	agatgggggt	tcacatggt	ggccaggctg	gtctcaaaact	cctgacctca
126361	ggtgatccac	ctgcctcggc	ctcacaaagt	gctgggatta	caggcataag	ccatggcgcc
126421	cggccagttc	aatgcttttt	aatatataac	agctttatgg	tgctgtgatt	accattccat
126481	acatctcagc	catctaaact	atatgctgaa	agtgaacaa	tcaaccataa	ccactattga
126541	ttttagaaca	ttttcatcac	ccccaaaaaga	aaccccgtag	acattggcag	tcactcccca
126601	tgttttcccc	aacctctcct	ccccaaacca	gccctgggca	accactaatc	tacattctgt
126661	ctctatgaat	tggcctattc	tggacatttc	atatacgtgg	aatcacatgc	tatgtggcct
126721	tttgtgactg	gcttccttca	tttagtatgt	tttcaagggt	catccatggt	gtagcatata
126781	gcagtatttc	attcctcttt	tattgtcaaa	taatacccca	tcacatgtat	ataccacatt
126841	ttatttatcc	atttgccggt	tgatggacat	gtgggttggt	ggtttttggc	tcctatgcat
126901	aatgctgcta	tgaacatctg	tggacaagtt	tttgtgtgga	catgtgtttt	cattcctctc
126961	aaagtggaa	tactgggtca	tacagtaact	gtgccagact	gtcctctaca	ggggctggac
127021	cattctcagt	gccccagca	gtgtgtgagg	gttcaatttc	ccctcatccc	caccaactact
127081	cattattctt	aggctctttg	actctagcta	tccagcagggt	gtgaagtgggt	gtcccatcat
127141	gcttcggctt	gcatttccct	gacaactaat	gatgctaagc	atcttttcat	gagcttatag
127201	gtcattttat	taactttgga	gaaacatcca	ttcagataact	ttccctattt	tttaataagac
127261	catttttctt	ttattgacct	gtaagagctc	tttatatat	ctagacacaa	atcccttatt
127321	agatatataa	tttatgaata	ttttctcttg	ttctgtgggc	tgtcttttca	actttcttat
127381	tagtgtccac	tgaagtcatg	aaggactttc	ttgaagtcca	gttcacctat	tttttcttt
127441	tgtccacttg	tgcttctggt	gtcatactca	agactttgcc	taatccaaag	taacaaagat
127501	ttactcctac	attatttaag	agttctatag	ttttagctct	tacatgtaga	tctatgatcc

127561	atccattttac	agtttaattatt	tgtgtatgggt	gtgaggaagg	gggtccaactt	cattttttcca
127621	tatgttgata	tccaggtgtg	ccaacacccat	ttgttgaaaa	gaccattctt	tccccactg
127681	acttgtcttg	gacctttgtc	aagaatcaat	atgaggatgt	atttccaaac	tctcagttct
127741	attctaccaa	tctatatgtc	tatccttctg	ccaacaccac	acagtcttga	ttactacagc
127801	tttgtagtaa	gtcttagaat	tacaaagcat	gagcccttcc	tctgttcttc	tcctcaatag
127861	catttttggt	agtctggatc	ccttgtattt	ctgtatgaat	tttaggatca	gcttgtcaat
127921	ttctgcacac	aagctagctg	ggatttgata	gaattaaatg	cattctggag	tcatgttttg
127981	taccatttag	ttcacctcct	aggccaatgt	ttaggaaaca	ttggcctaca	aagaaactct
128041	gcattagaaa	ccaagggctt	ggtcaggcat	ggtggctcac	accctgtaat	cccagcactt
128101	tgggaggctg	aggcaggcgg	atcacgagg	caggagattg	agaccatcct	ggctaacacg
128161	gtaaaacct	gtcatcttga	gatgggagcc	ctaagtctat	caactaacca	ggtgagccac
128221	tgtgagcaac	tctggggcct	gggagagggt	gcacagaggg	atggcttcaa	ccctgctgct
128281	agagaatgcc	ctcgagagcg	tctgtctcag	ttctgcttgg	catccctccc	cagcaccccc
128341	cacctgccac	tttacgcagc	ctggacttga	ctttaaaact	gagacacctg	gcaagtgggg
128401	aaatcacagg	tcgaaatagc	ggaattgccc	tctgcatcta	tgctacaggg	ataatgaggg
128461	ttacaaatca	gaagacattt	agcatatatt	ctggaaaaaa	aaagctatgt	aatatataaa
128521	ctattcataa	tttaacttta	ataaatgggg	gcaggcaaaa	aattgccaaa	caagcccaca
128581	acaatcagac	aaagagtaga	ggaagaaaaa	aggaattaca	gtttaattgc	actcgtttcc
128641	taatcttcag	aatcttcagt	tagcatcaac	agttactggt	aagtttctaa	aatcagtaga
128701	tacaagctga	agaattctac	tcaaaattag	aaaagtcaca	actagtaaaa	ctaattgtata
128761	ccttccaaat	gactgaaggg	agaccatata	gaaaaatagg	tagactagaa	aaacagcaaa
128821	aaaaaaaaaa	aaaaaaaaaa	tccagaaaaa	cagaaaaacat	aaattagatg	acagaagtaa
128881	gaacaaatat	ataacaataa	atgtaggctg	ggcatgatgt	gatcccagaa	ctctgggagg
128941	ctgaggcagt	taggaatttg	agaccaggct	ggccaacatg	gtgaaacccc	atctgtccta
129001	aaaatacaaa	attagccaga	cgtggtgggt	cacacctgta	atcccagcta	ctcaggaggc
129061	tgaggcggga	gaatcacatt	aacccaggag	gcagagttac	agtgaagccg	gatcatgtca
129121	ctgcactcca	gcttgggaga	cagagacaga	ccttgactca	aaaaaaacca	aacctaaaaa
129181	taaacaaaac	aaacaaaaca	ataaatgggt	agattattat	tattattatt	attttgagga
129241	ctccagcctg	ggagacagag	agataccctg	tctcaaaaaa	acccaaaatc	caaccgggag
129301	cagtggctca	cgtccataat	cccagcactt	tgggaggagc	aggcgaatgg	atcacaaggc
129361	caggagtcca	agaccagcct	ggccaattat	gtgaaacccc	gtctctacta	aaaacccaaa
129421	aattagccgg	gcttgggtgt	acgcgcctgt	gggtccagct	gcttgggagg	ctgaggcagg
129481	agaattgctt	gaacccagga	ggctgaggtt	gcagtgaagc	gagattgcat	cactgcactc
129541	cattctgggc	aacagagtga	gactccatct	caaaaaacaa	caaacaaaca	aaaacagaca
129601	acaaaaacaa	taaatgggtc	gattattatt	attattattt	tggagacagt	gtcttgttct
129661	gtcacccagg	ctggagtggg	gtacagtggc	ctgagctcag	ttcactgaga	cctctgcctc
129721	ccagggtcaa	gcaattcttc	tgccctcagc	tcccaagtag	ctgggattac	agggtgcaca
129781	caccatgcct	ggctaatttt	tgcattttta	gtagagacgg	ggtttcgcca	tgttgggcag
129841	gctgggtctg	aactcctgac	ctcagggtgat	ccgcccacct	cggcctccca	aagtgtctgg
129901	attacaggtg	tgaaccaccg	ttcccagcca	gtcagattat	ttcaataaca	ctctatgccc
129961	tgcaaaataa	ttttacaaaa	tttacagatt	tatgctaggg	taaaagagtt	tctcagattt
130021	tttgaccact	taagggaacag	acacacccat	gacatattgg	tgaccgaaga	aggcggcaag
130081	cccagctgca	gcatgacctc	cttacgtaaa	atcatgtata	tctctgtgcc	aaacgtcagc
130141	acagtctaat	aattaagcat	gtggcctccg	agtcaactct	cctcgattca	aatgtttact
130201	tcacctctga	ctgagtggcc	tctctgtgga	cagtgtctgc	agggttgaat	ggagataata
130261	acagggtttc	catggctgat	gtgagggcta	agcgagatac	cgcgtctcta	gcatgggca
130321	cagcacgtga	taaatcagcc	gactggcagc	accatcacca	tcatgatgat	catggcatga
130381	cagctgttag	gaaacgcgta	acaaagggtc	agcagtgtta	cgatcggggg	tgagtttcac
130441	tttttattaa	tattttctgt	attttttttg	ttttacattt	ttacaaaggg	catatagcaa
130501	aacagacaat	tcatgtgaag	acaggagaga	cggaaagggg	gagagcacag	agagaaagat
130561	aagggggaaa	atggaataaa	caagatgccg	acaagaatgc	ctcacggggc	cagcgcaggg
130621	gccagccagg	cggccttctg	ctaaatcaag	caacgcttga	gtggctccct	cgggacctgg
130681	cctgtccaga	atgtcccaca	ctcagccctc	tgacgcccag	tccccatct	ccatggctcc
130741	tgacacaccc	aatgccgggg	tagcctccag	gaaacagagc	tggccagcca	gacccagcca
130801	ggacacagag	gtcagtgggt	cagacccaga	ctctgggact	gcaacaggcc	tgccattttg
130861	tgacgagaac	tgacactcgt	cccggccaga	ttctctggct	gacatttccc	agggtctcgt
130921	cagcatgaag	cccacgtcgg	ggaacgtccc	cagtcacaaa	ttaacattca	cataaataac
130981	agggcgctct	ctactaaaga	agaaaaatga	ctcgtgctgc	aggaacattt	gccaaactat
131041	caaaaagcaa	gatttttctt	tctgagcttc	catgtctcac	cagctggacc	acagcagatt
131101	gtggccttga	gacgtgtggg	ggaggagagc	tcattggaata	agggtccaact	tccccattca
131161	agatcctgcc	aacaggccgg	gcacgggtgc	tcacgcctat	aatcctagca	ctttgggagg
131221	ccgaggcagg	cggatcacct	gagggtcagga	gtttgagacc	atcctggggc	acatgggtgaa
131281	accctgtctc	taccaaataa	caacaaatta	gcgggacatg	gtggtggggc	tctgtagtcc
131341	cagctactcg	ggaggctgag	gcattgagaat	cgcttgagcc	cagaaggcgg	agggtgcaat

131401	gagctaagat	cgcaccgctg	cactccagct	tgggctacag	agtgagagac	tcttgtctca
131461	aaaaaaaaaa	aaaaaaaaaa	agaccctgtc	gacaggccct	tccctgggct	ggcacctgcc
131521	cacaccagct	ggaacctaac	acctttgcac	agctgttccc	tgccagcctc	cctgctctct
131581	tggaggtctg	cacacaccgc	ccctcccttc	ctccaaccag	cccgggagac	cagcacgcct
131641	gagcaccagc	actcccagag	acagagccgg	cacgggtggg	ctgtctgtct	agggggctgg
131701	ggtggcaggg	gcacatgtct	gggagttgcc	aacatatgct	tccttaccag	attccccagg
131761	ggacccaccc	tggagagcaa	ccaaaccctg	caaggctgca	ctccaagaac	tgagggcctg
131821	gggacccagg	accgtccccc	agcattcaca	gcaggctcac	ccacatccac	caggagatgc
131881	tgcccaggta	gactgccagc	accccctccc	acaaaaggct	ggggcccgcg	ccttccctag
131941	gttgctcggg	caaccatcac	agggcagcac	aggccggggc	ttgaatagca	gaacctattc
132001	tctcacagtc	ctggaggcta	gagctctaaa	attaagggtg	gggcagaggc	ggtttctctc
132061	gagccctctc	tccttggctt	gtcgacagca	tctttccctc	gtcctcacat	ggccgtcccc
132121	ctgtgtgtac	ttctgtccta	atcgcttctt	cttaggacac	cagtcattca	tgttgaatca
132181	gtgacttcat	tttcagttaa	ttatctcttt	aaagaccctg	cctccaaata	cagccaatat
132241	cctgatctac	gggagtttag	gacctcaaca	aaggaaactg	gggatctagg	gagggcgtca
132301	ctcagccctc	aacagccctc	aggcgccaag	tcacacacag	atcggagcag	gctgtctggc
132361	gggtccgggtg	gccagagctg	ggcgggctgt	tccttgggga	gccactctgg	acaggccgga
132421	aggtcaatgc	caggatcact	ccccaggggc	cagaaatcca	gcatggagga	aagctgccac
132481	tccaggggaag	taggagcccg	cccagggctg	gccacatggg	atgcccatct	cacccttctc
132541	tctaaactgg	gcattctcac	atggggacac	ggggcagtga	gtgcaggcgc	agggagacga
132601	gaatgtcgca	cagccggtag	gtggtggacg	ggcctcagga	cccctccgcc	agcagcagge
132661	cactctctca	cggtcacgca	ggagccctgc	tccagagccc	acgactggag	ccagctgagg
132721	ctggtgtctc	acattcctgc	agagaggctg	gctaaggctc	ccgtgtgggg	gcactcccca
132781	tgccactctg	accccagaat	tttctgggcg	atcccaaggc	tgagtctacg	cagctgttgg
132841	aagacggggg	atgcaattct	agatgtgtgg	cgtgttttga	acagcatcca	ataccctttc
132901	tgtggcagct	cagcccggag	agcaggtccc	cagtctcgag	gacccaggag	gcagggtctg
132961	ctccttccca	cagcgggtgc	tcctccagga	ggggccagg	gctcagaggg	aaaccagtgc
133021	ccaccaaggg	cccagctgcc	tggctggctc	ccagcagtgt	ggggaagaca	cactaccgga
133081	cggacgtccc	ccttagaagc	aggggcagac	aggaactccc	atggggagag	cacaggctgg
133141	gacgcagccc	gcccacgcgc	agtcccccca	ctccactgtg	caacaaggat	cagcctttcc
133201	tctctcaaaag	ctcgggttcc	cccagagggg	gctgcagtgt	agagttgatg	gtgagactca
133261	gtggggcatg	tcctccatgt	ctgcttgtgc	aaagctgccc	agatcaggca	gttttccag
133321	ccagaggacc	ccccctttca	ctttccatca	caggctgtca	tctccccggg	tccttggggc
133381	acaggccctt	ctcaggaccc	ctccggcctc	atgcacacac	agagcctgtg	cactgcctca
133441	ctcctgtctc	ctttgaattg	ggcttctttt	cctgtctgtg	tctctggggc	ccactgtgct
133501	atccagggca	cgctgaccgc	cgacgttctt	ctgcccgcct	gccagctgc	cactccgggg
133561	cgggcagccc	tgccaggccc	gaggctgtct	gcttcccctt	ctgtccagag	ttcccaaaac
133621	aaaaatcagg	ccataccatc	tgacgaagga	tcctgtctgg	tttatttatg	ggaagagcca
133681	gggctctgaq	ggccagggtg	tctgggctgt	ggccaggctg	ctctgcaggc	cttcagacac
133741	tgcccagccc	gtgtccccac	ccaccagggt	tggctccggg	ggagggggcg	gggcagggcc
133801	ccagcactgg	tgagcccaca	ctggtgactc	tgggatgtgt	atltattcta	gagccaggac
133861	agcagccgcc	agggccctag	gcccctctgt	gggtggttgg	cccactggag	agccggagca
133921	gcgaggccag	caatgaccaa	aggcgccgta	tccagaatgc	aggcaggctc	aggcgcacct
133981	gcccctccctg	gatgccacta	cagattcggg	gagactgaac	tcattgggatc	aagtgcacca
134041	cagatcctgg	tcctcacaag	ggacaaggcc	accaggtgtc	acactgggag	acagtggcac
134101	cccagggtat	gaaaggccac	accaggctta	tgtccagtaa	ccctgcctgc	acagcacagc
134161	acggcacggc	atccataccg	cagcacagcc	ccgcctcccg	gcagtgccaa	gcccgggac
134221	tcaggggccc	caccagaaac	cagacagggt	agcaggtggc	agctccaggc	caggtgctga
134281	ccttgtgcat	accaggcctt	ccccgagagc	aggctgcagc	ccccagctg	gttctgactg
134341	ccacaggact	gcgcccctcc	tgccaggctc	tgctcacacc	cacactccct	gctcagagac
134401	ggcctgtccc	accgcccctg	gtgaaatcac	acccaagcac	agcggaaacg	tggaggaggc
134461	agctggaagg	tgctttcggg	ggtcaaatct	ggggcaattt	gaacacaata	ataaatgata
134521	gtaataggtt	agaacccatt	gaataaaata	aatatccaca	gatcagtatt	ggtagaaaca
134581	aataactgaa	taataaataa	gtgggagggg	aagggcagct	ctttctcaca	gggtaattca
134641	actaaaaatg	taaaagaaat	gatgagatag	aaaatcatgg	tagtaaaacc	catcggaac
134701	agcagagtca	ggactgcggc	aggtacgact	tatccgtgga	ggctacaatt	agtgagcaaa
134761	aggataacaa	gaaacagaat	gtcacatca	cgtcacagca	gctaccacag	ggagatttaa
134821	tcattcgcaat	gggaaaaaag	agtgaagag	tggtttgctc	atcaagccac	tcattctgaga
134881	cctgaaaaaa	cctcacagtc	acagccctag	agctcccggg	gcattccagt	gaatcccagt
134941	aggtaacacc	ttagccaaat	gaagaagggc	atcactgaca	ataggagaca	actggacatc
135001	ccaatgtgac	agtcccagcg	tggtgcccta	cggagggcac	agtgtctctg	tgacagctct
135061	gccagaaaca	cgtaagctca	agccagtaca	aggcacacca	gaaaaaccca	tggggatgca
135121	ccgtctacaa	aacagctgac	cagtcctctt	caaaaacatc	caggtcatga	aacctcagat
135181	gggaggaaac	gggagagaca	caacaactaa	atgccacgtg	ggatcctgag	ggggagcctg

135241	aaccagacaa	cggacatgag	tccggaagact	gacaaaattc	agtgcagtct	ggagtgcctac
135301	ggttagaatg	tgactgggtt	gttccaccac	aaactcatct	tgacatttgg	tccccagtgt
135361	ggcagtggtg	gaaggtgggg	cctgggggtg	ggtatctgaa	tcattggaggc	cgatccctcg
135421	tgagtggcct	gatactgttc	tcaaggtagc	aagggaattc	ttgctctgga	gaggcttgat
135481	tagttcccg	gcctgggttg	ttataaaagc	aggacactcc	ctgggtttcc	cttctttgca
135541	cacatgggct	ttccctttga	ccctctccac	cacgttatga	cacagcacaa	aaaggcctca
135601	ccagaagcca	gggccatgct	cttgaacttc	ccagcctgca	gaaccatgag	ctaaaataagc
135661	ttcttttctt	tataaattac	ccagtctcag	gtattccatt	atagcaacac	aaaatagact
135721	aaaactaaga	cactaagtac	caacgcttag	tatcaacct	caaagttcct	ggtttcagtc
135781	cttgttctac	gttgcataca	gaggaagctg	ggtgaagggc	acagaggaaa	tctatgctat
135841	ttttgcaact	ttctttgcaa	tttttccaag	acagcataga	tctaaagtta	tttcaaaatt
135901	aaaactttta	acacataaaa	ataaaaactgg	gccccagcct	ccctcagtgc	tctctggccg
135961	gtccattctt	cacattggct	gcctttccct	ggagtggagg	cgccactgc	atccccagtg
136021	tccagggagc	aggtcctgta	gaaggtggta	ggttccaggg	atcggggagc	aagtacagag
136081	gccctgagaa	ggaagcaagc	ctggcgtggt	ccaggaacac	cagggagaa	cgtgtagccg
136141	gagcatagct	cttgtgcaga	gggacctatg	ctgagccgcc	ctgtgggcc	tgccgttccc
136201	ttcagggcac	tctccaaccc	tctgccctgt	gcatggggcc	acggatgccc	caggagctct
136261	agggctgttg	actatgaggt	tgtctcaggt	ctcagtgagt	ggcttgatga	gcgagccacc
136321	ctcactggcc	aggaggggac	ttccttcaca	ggtgtgaatg	ggactgagcc	accagcatga
136381	tgctgacacg	tgggttttga	ggattccac	tcacatcaga	cagaatgagt	gcggggctct
136441	gggcatagac	tgcccaagcc	tcaagcctga	gtggcaactg	tgggctcctg	catcggctgt
136501	ggggccacct	ccgggctctg	gtgttccagg	gtgggcagtg	gcagcagaca	tccttctact
136561	tgtgcctctg	tcactggtgc	ccagggcgtg	ggccacacac	ctcaaaggtt	ccgtgtcatt
136621	tccgggcagt	cttccagggc	tggctgcac	cacctgcttg	gagtgtcttc	agagggaagc
136681	tgttgccttg	caatgccacg	acccttcccg	tgtcctctct	tgccctccag	ccgtgtctga
136741	tgttgccttg	tcctccagg	ctgggtctga	gagggatgga	atgcctgtca	caactgggat
136801	ccctcttctc	ttgtcattct	gcctgtctgt	ggcagagtgt	ggaaggcggg	agtgtgccc
136861	ccattaaacg	aggttgaatt	ctgggagtg	agaagtaaca	gcctcaggac	tgatggggaa
136921	gatgaggctg	caagatgtgt	ggttcttccc	ttcgctataa	aggtccaaac	gatatttatt
136981	tattttattt	tttattttag	gacaggatct	tactgtgtca	cccactgcag	tacagtggct
137041	caaacacagc	tctactgcag	ctggagcccc	tgggctcaag	tgatcctccc	aactcagcct
137101	cccaagtagc	tgggaccaca	gacgtgtgcc	acaacagcta	gctaattaaa	agaaaaaaa
137161	aaatttgtag	aaatgaggtc	ttgctatggt	gccagggtcg	atgttgaact	cctgggctca
137221	agtaatccta	ctgccttggg	ctcccaaaat	gctgggatta	cagatgtgag	ccaccacacc
137281	tggcccaaac	attaaattta	acacgaccag	atgaggagct	atggctcaca	ttttcttgca
137341	cactctgggc	aagatccaca	ttggtgggtc	tcattaatga	agcacggtaa	ccaccttcaa
137401	gggaacaacg	gaagccggca	ttcccggggg	cagagagaag	cggcggagtc	aggagaaatg
137461	ccacaaccag	cactaaatac	atggggctgg	ggaatgcggg	gcaagaagat	atacagcatc
137521	tccacacaaa	caggttccaa	agagccacac	attgtgagcg	gtaacagctt	tgctattacc
137581	ccccagctgt	gggagatata	aatatgaaat	tagctgggac	agcacactcg	accataaaga
137641	acaggaagag	atagctcctg	agcctgagca	gtaacgaggg	tctctgcaag	cactgagcag
137701	acagttagca	aatgtgttcg	tgtctctcct	ccccgtcccc	taaatcttcc	ttgatatact
137761	ttgctattcc	ccatgtgggc	tcatgaatct	tgttataaat	tttttaatac	aattgtctat
137821	atagacagga	gtcctgcagg	aaatataatg	ccagggtccc	atgtatccta	gggacagcca
137881	gccatgcctt	tgggaacagc	ctagaaggac	tgtcaggaag	gggcctgggt	gggtgggagt
137941	cagcttgccc	aaggtgactc	tctctccttg	caagatacat	ctgagagaga	agcccattgt
138001	aagcagggcc	aatccaagct	cccaccccac	cctggagtgc	agaccactgc	ccagacggct
138061	tcctgaggtc	ctggctgggg	ccctccaagg	tacccattt	ccaggcctcc	agccccagcc
138121	ccccaatagc	ctgccttgaa	ttgtcttctg	cttaagttag	ttgaggtcag	ttttgccatc
138181	caaagagaag	ccctcctggt	ttcctagagg	ttgtcaggag	ctacgtctgc	cggctccctg
138241	ggcttgggac	atcctctggt	gttgggggga	gaggggtggg	agcatcttct	ctgcagccat
138301	gacctcacct	tgggtgacag	cacagcttct	cactcacagc	actggtgtaa	gtgctaacct
138361	gaagccgctc	ctccttgtcc	acgggggaaga	cagtctgacc	ccactttcct	ctgctccccg
138421	aaaataagga	acgttatttg	tgttgctttg	tcctgacaga	ctcacacgtc	cttaaaaaat
138481	gctaactctt	acaaaagagg	cagcggcctg	tgcttccctt	gaggaaaagtc	agattaatta
138541	gctgactccc	acatgagcag	cctctgattc	ttgcaggctt	tgtttatgga	tcagaaatca
138601	ggcttgaagt	gaaaactgac	atctcaatgt	gaagagtatt	ttccctgcta	actgcagtga
138661	gtcagagccg	tgcgtgagc	tagacagata	gggaacatgg	ccttcccatg	tggactggag
138721	tgtgccaacg	cctccggggc	acagcccagg	acaagcgtgg	gacatcagcc	tcccattgtc
138781	agggcagaac	agggcagtcg	cccatagatc	ccagcccaag	acaagcaccg	gagttggcat
138841	tccagggtag	ggttttgggt	gtttgtttgt	ttttctcaag	atggaaatctt	gctctgtcgc
138901	ctaggctgga	gtgcagtggc	atgatctggg	ctcaccacaa	cctccacctc	ctgggttcaa
138961	gtgattctcg	tgtctcagcc	tcccagtagt	ctgtgattac	aggcacacgc	caccacaccc
139021	agctaaatct	tttttttttt	tgtatttttt	tagtagaatg	gggtttcatc	acattggcta

139081	ggtgggtctcg	aactcctgac	cctgtgatcc	gcttcccaaa	atgctgagat
139141	tacaggtgtg	agccaccgca	cctggccaag	gtttttttgt	ttttttttta
139201	aggtatattgt	gcatctcaaa	cattccaactt	aatgactggg	aactggggcc
139261	gcacttttgg	gaggccgagg	caggaagact	gcttgagccc	aggagttcca
139321	gtgaggcctc	atctctacaa	ataatttaaa	aattagctgg	gtgtgggtgt
139381	ggtcccagct	acttgggagg	ctgaggtggg	agaatcacct	aaaccctgaa
139441	gcagtggagct	aagatgctgc	cactgcactc	cagcctgggt	aacagaaaaa
139501	aaaagaaaaa	atatgactgg	gaaccgggag	aaagatgttg	atgaaggcac
139561	gtcagagctg	acagcccaga	aacacgtcca	gccagccaat	tatttttaat
139621	cctttgggtga	ctcaatccca	tgaacagggg	agatgggtgc	atggctagaa
139681	ggctccgagg	gctgcccagc	agaccacccc	taggggggtgc	caggatcaac
139741	gatgctgagc	acccatgcaa	gtctgaccca	cccacgtggc	ccaaggaggc
139801	ggtccggaat	ggtggtctgg	agattgggca	cccacagact	tggggagtgg
139861	ccaagccctg	ccccgcctcc	cactgcagct	tccccgcagg	gctgcaccca
139921	gtggcccctg	ccccgtgtgc	ccagcttcca	aggtggcctg	cccctgatgc
139981	tcttgctggc	caagcggagg	ggcttgcttt	ctagggctag	caggaaatgt
140041	cctgaaaaat	ccttcaattc	cctcattgag	tctccagtag	acttagtca
140101	caacaggaca	actcttttcc	accccaaagt	ctgacagcct	ctgaggcagt
140161	aaggaagagt	cactgtgcag	actcctggca	tggaggccct	cctgtcgctc
140221	cctctaggcg	aacgggcctg	tggggaactt	ggcgggttcc	cactggccgc
140281	gtttccactt	caaaaccagg	ccttgagcac	cctactggag	gctgctgggt
140341	tctctagagc	ctgagagcat	ctggcaccag	gcctggctga	ataaatagat
140401	aaaggaaacc	tggctttgaa	aagacacaga	gaagaaccag	gtccctgcgt
140461	caggctgatg	ttctgggctg	gcggtgtctt	gttgggtctc	ccacccttcc
140521	ctgtcctaca	gcccagtagg	cagacctgat	ggactggctc	ctcctggctc
140581	aaggggggca	ccgacaggaa	gccagggctg	gggagagaga	cagaaggttt
140641	gtgtgtgggt	ccccagcctt	ccccgcctcg	caagctccct	cgctcctgcc
140701	gagttgtgcc	tttgtcaact	cctttctgtt	gatecccttat	gtgtcccagg
140761	atccgacatg	ccctgtagcc	cgaccaggca	gccctcactc	tcatagtcca
140821	ccatctctag	ggtttgggat	gggacttcca	cactccagcc	caatttccac
140881	acgggtgagc	cagttccctg	caaccaccac	agactcagga	cacaaagtcc
140941	cagaaattcc	tgagcagggc	agaggagccc	cgtccatggc	gggtggggca
141001	tgagagtcgg	ggtgggagga	gctggaggct	ctggaaggga	ttcaggagca
141061	ccaccttcca	cacacgtcct	caaaagccct	tccatatccc	gctgaaacca
141121	ctgtgctggg	agcaaggagg	gcgtggagac	aggtggggct	gtggaaggag
141181	ggaggggaag	ctcctccaac	ccaggacgcc	ccactgcagg	aggaaatgcc
141241	tggtccaggga	gactgagggg	cctgctcagc	cacacagcac	gtgccaaagt
141301	ggggaagagc	aaaccgagga	gctggcgggt	cctcggacga	tgcagtgatg
141361	aggatggggg	aggcctgagc	cgtggggaca	ggagaatgac	cacaggatcc
141421	ggtccgtggg	acctcagcag	acgctttggg	tgatgccaga	gagccatgga
141481	gcaggggctg	tgcttctgag	aatgttgaca	atgatgggac	ccgggagcag
141541	agggaggggc	aggggcaatg	ccggggcctg	tctgtgctcg	gaggaggggg
141601	gcaaggaggt	gggtgagcag	ggagcgggaa	ggtcccaagg	ggggaatgct
141661	ccagcagtcg	ggaatgtcga	agtatagcct	tcagcgaggc	tggggcccct
141721	aggaaagtaa	gggaggaggg	atggggctgg	accagcaaat	tcagcaaggga
141781	cgggagacag	tggtacaggc	agggagcaag	tacagcaaag	aggaccagtg
141841	cggataagga	aggaccgcga	ggcaggtgga	gatgggctca	gtgggtaagg
141901	cggcctgggc	agagggcagg	agggagaggc	ggcgggcaga	gagaggctgc
141961	gggtcaggag	ggtggtgcct	gggaagacca	gctgcctgcc	tgctcccagg
142021	gacaagtcaa	gcataaagat	gcaaagggtg	aggccgaaaa	aaggacgggc
142081	actaacacaa	ccatcaatgg	gtttaacatt	gcagttagag	atcaaccaga
142141	acaaacaagg	catgccgtct	gtaacacatg	cctgaaacat	gctaaaaagag
142201	aactgcaaac	ccagaaagca	gggtgtctgt	agtatcatca	aaatagaact
142261	actaaatgac	tggaaactcc	tgttgtattg	acaaggcttg	gtgaaagcac
142321	agcttttaca	accaagggcc	gcagcatctg	ccatggttag	aaacacaagt
142381	ccctgtcaaa	gggaaatgcg	aacggccact	cccagtgggt	gacaggtctc
142441	gcagaggggc	acagactcat	tgactcggga	aatgggcccg	attgatttct
142501	gacgacttac	aaccgaaagc	aaatgagtat	tctttacaag	catgggggag
142561	aattgatccc	ggcctcccct	gcctgcctgg	cctgaatccc	ccaagcctgt
142621	ctcccacccc	acctgcctgg	caatgcagaa	acctgagatt	gctgcttctc
142681	accaagaggt	gaagttggtg	tggggacccc	cttctgcctc	actccctgct
142741	agccagaagg	ccctccccac	agcctgcccc	tttctcttga	ctgttccctt
142801	ctgaggcccc	atgtctccca	ggaccctccc	tcaagacccc	ggcttccctc
142861	cttgggcccct	tgcattcatg	ggcataatgc	caactcccca	cccctactga

142921	ggcagagtgc	ccccaacgcc	ctccactata	atcccagagg	acggagagct	cttgctctga
142981	gccgagacaa	cccactgccc	accaccagct	ccctcagggc	tctccacaag	tcccttctcg
143041	cctcccctgc	tgcgtcttcg	gcctcctcct	ctctaaagtc	ctctctcagc	ccaacaacag
143101	aaccgggtct	cttcaatcct	gaacaagaag	actctgcttc	ccctccagcc	acaggacctc
143161	ccttccctca	cagcccttct	cagcaggctg	gcctctgctg	ctccctccca	cgcaccaatg
143221	cccatggcct	ctgtccctcc	tccccagggc	cccggggacc	tccacaccct	cctgctcagc
143281	ctctctctgg	ccgccccag	ctgctccggg	tccacagca	ccagaggcca	gggtccctc
143341	gtgccacagt	gacctgcaa	cgcagattcc	gtgctccctt	cagaagccca	cctgacattc
143401	atagaagctg	agcaaacacc	aggagacaga	ggcaatttca	gaaattcccc	gaggctcggg
143461	gaacccaggc	tgcccccttc	actcctaate	tgtggcaggc	ctggccctgc	caagcctcca
143521	tggagatgag	ctcttctgt	cccaagggtg	tgggtggaag	cgggcctttg	gaaggctgat
143581	aggctcatgag	ggtggggacc	ccgttaatgg	gattagtgcc	cttatcagag	aggcccaggg
143641	agcttgtgca	ccctccactg	tgtgcaaaca	cagcaagaag	gccccagcta	tgagcgagga
143701	agcagccctc	accagacacc	aaacccactg	gcgccccgat	ctcacatgtg	agactccaaa
143761	actgtgagaa	ataaatctt	gttgtttcta	aacaccaggt	ctgtgatatt	ttgttatggc
143821	agcctgaacg	gaggacacc	tgacagcct	gggatgtcga	gaatacacac	gtggcaggag
143881	gggtgtggac	gggctctacg	gaactgggag	atgcctagaa	atgctagagg	gggtgagggg
143941	gaggcattca	caagcccttg	taggaaactg	agatctgcgc	tgggttagggg	tggggtgtca
144001	ctggaatggc	aagatgcaga	tataaatgct	gacaagacc	aatttaaacc	cagagatgct
144061	aaggtctgaa	aacatttggg	ttatactagg	aatcaacgca	aagcttaata	acctaaaatc
144121	caaccactcg	gaagtttttt	caacattcct	ctaaatagct	cttaagtta	tggaaatcaa
144181	atcataaatc	tttttaaat	gatgcagtg	agaattcagc	acatcagaac	ctgtggatga
144241	gaccaaagcc	accaccagaa	gcaagttcac	aaccttctta	attatagacc	agagtgaaaa
144301	taaatgaagc	cagtgtctca	gtcaagaggg	cagaagagag	agcaaccata	aacctaaaga
144361	tgttgataca	aatagcatga	gtgacaagat	ctcacttctg	agtgtgacat	atatgtaaca
144421	agatggaaaa	acggttagtt	tgtatagctg	taacaaatta	ctgcaaatgt	agtggcttaa
144481	aaacacaaat	ttgtccaggc	gcggtggctc	acgcctgtaa	tcccagcact	ttgggaggcc
144541	aaggtgggtg	aatcacgagg	tcaggagatc	gagaccatcc	tggctaacac	ggtgaaaccc
144601	catctctact	aaaaatacaa	aaaattagcc	aggcgtggtg	gcgggcgcct	gtagtcccag
144661	ctacttggga	ggctgaggca	ggagaatggt	gtgaaccag	gaggcggagc	ttgcagttag
144721	ccaagatcac	accacagcac	tccagcctgg	gcaacagatc	aagactccaa	ctcaaacaca
144781	cacacacaca	cacacacaca	cacacacaca	cacacacaca	catttattat	ctgacagttc
144841	tagagatctg	aagtccacaa	tgggtctcac	tgggctaaaa	tcaaggtgtg	gggagggctg
144901	cattccttct	ggaagctcta	aggagaatgt	ttccttgttt	ttttccagct	tttagaggct
144961	acccacattc	cttggtctgt	ggccccctcc	tccatcttca	aagccagcac	cggagcatct
145021	actgacccct	ctctgactct	gaacttctgc	ctccatcttc	tacttttatg	aatcttgtc
145081	atggccagct	cacacctgta	atcccagcac	tttgggaggc	tgaggcgggc	ggatcacaa
145141	gtcaggagat	ggagaccatc	ctggctaaca	cagtgaacc	ccgtctctac	taaaaataca
145201	aaaaattagc	caggcgtggt	ggcaggcgcc	tgtagtccca	gctactctgg	aggctgaggc
145261	aggagaatca	cttggaacca	ggagggtggg	cttgcaatga	gcctgagatt	atgccactgc
145321	actcagcct	ggcgacaga	gtgagactcc	gtctcaaaaa	aaaaaaaaaa	aaaaaaaaaa
145381	gaatccttgc	cattacactg	tgccatccag	gataatctcc	ttatcttaaa	agtcaactaa
145441	ttagcaacct	gacttccatc	tgtaaccttc	actgcccctg	ctatgtcacc	tgattttatc
145501	ccaggttctc	gggactggga	catggacaac	tttgggagaa	ggattattct	gtctctcaca
145561	gagactatct	caaaaaggga	ggatttcatg	gtttttgctt	tctccttaat	acttttctac
145621	attattcaat	gataatgacc	atatgttttg	tatgttcaaa	agtaattttt	aattgtgaat
145681	aatgtaaaat	attgtgaatc	atgttttgtc	agttgtttgt	aataatgttt	aattgtaaac
145741	aatggaaaac	taaagaaaca	aaagaatagc	atcctcttat	agcaccatt	gatgtctttg
145801	caacactttt	tggctgggtc	tttgtccccg	tccccctgac	tttgagctgc	ctgaggacaa
145861	gccagacccc	ccttcaactc	tgtggctcta	gcagtaagca	tgaggacttt	ggcacagcag
145921	gcttttagcag	acatttctca	aagtgaagat	aaatccccta	ataaaggagg	ccctggttta
145981	agaagaatga	tgtgcacact	caaaaagtga	ctttaaggcc	aagctccgaa	taaacacgat
146041	gatattatct	tatttaagag	tcttcaatac	tacaacttaa	ttcagtaaca	gtctgagaaa
146101	cgcaagccag	cagacagata	aaccgttatg	gctctggctc	tggcaaaaa	tcttcatcta
146161	cgcgaggcca	gaagaaagcc	aattaaaagt	cggagaggtc	ccgcaggaca	gaggccggag
146221	gatgtcccca	cacactgagc	acaagcccca	aacacaggcc	tccccagtct	ggagccggcg
146281	gccagccccc	tcctggcctc	agtgaagcct	ggactcctca	ctctgtccaa	gtgggaggaa
146341	ctgggtccag	ctgcaagaac	cgcctactgt	gatctgggca	aaggttctgg	aggctctgct
146401	cgctcctgtc	ccccaggaaa	aaaaaaaaaa	aaaagacagc	agagggcatt	gctcaggggc
146461	ccagcctggc	tgtcatccct	gaacatcctg	ggctgaagga	aagaattgag	cccagccgaa
146521	acttgagatc	tttagactta	atccccactg	ctgtcacttg	tggctatttg	gcctctttcg
146581	acctctgctt	ttccgtgtct	aaaatggcac	aatattacct	actttacaga	gttgtcaagg
146641	gattgaaaat	aagataatga	aggcaaagga	agttagctca	gtatcattat	gcccatttta
146701	cagattggga	aacaggctcc	aagagatgac	atgactcttt	caggattaga	cagctaataga

146761	gtgggtggatc	cggatctgct	cgcacagggc	tgggctcctc	cccagccccg	cactgggaca
146821	ggggcatctc	ttgagggctg	ccctgtggcg	cccagatgcc	gcagggagac	aaggctgtgt
146881	gcacagatct	ctctccactg	ggcagcacct	gagtaactgg	gggagagaag	tgggtttcaa
146941	gcatctcctt	tgtgaaccac	tgggcttaag	tatgctgaac	ttcagttttg	acgcagaact
147001	tggagacaac	ccatcctctg	ggaaaagaca	caaataaggg	caggacccct	gcaccatgtg
147061	ccctgctgcc	ccagcagtg	cctctcatcc	agcgtgtgaa	gaaggagtgc	agccggagag
147121	ccggtggcag	cagggcgctg	gggatgcccc	gctcacagtg	tgctgtgcct	gggagaggag
147181	cgagggcagg	ccccagctg	cttgggaaca	gtcagcctgc	acacctcccc	cctgggtgcc
147241	cttccaaggc	agccctcagg	ccccaggtc	cgggactccc	tgacagggaa	ctgcaagcca
147301	gtcctggccc	agcacagccc	ttccgccagc	ctctggccag	cgtcagggaag	gggagtgtag
147361	agtgagttag	aagccacaca	ggagcagaag	ccgggacttc	tgctctcac	cccactgc
147421	gtatgaaaat	agacgggggt	gatggccggg	tgccgtggct	catgcctgta	atccagcat
147481	tttgggaggc	tgaggtgggc	ggatcatgag	gtcaggagat	ggagaccatc	ctggctaaca
147541	cagtgaacc	ccgtctctac	taaaaaata	taaaacatta	gccgggctgt	gtggggggcg
147601	cctgtagtcc	cagctgctct	gggcgacaga	gcaagaccgt	ctcaaaaaaa	aaaaaaaaaa
147661	aaaaaaaaaa	aagcggggga	gtgataaggc	cccggacaac	ccaaactgca	gcaagattct
147721	ctgcggaggc	agagctgcct	cctcctgggg	gtcctctgga	ccaggaagga	gccccaggga
147781	cttgtggccc	agtgggtccg	gcttcccaac	ctccaccttg	ccagcagcac	cgatagatgt
147841	ggccacaaa	gtgcattcac	atttgcgacc	tgtggctgaa	tgacagacct	gggtgctatc
147901	tcccgggggg	gcaaatggca	gtaagcaagc	aaagaggaag	tcgacagaga	gtgacgcagc
147961	cgtgaggaca	acagcacaga	gtattgttga	ggggggcggg	ggacttctct	gaagagatga
148021	cactggagct	gaggcctaaa	tgatttgaaa	gaggcagcga	tgagagagatc	tggggaagt
148081	ggctccagca	cagagaacag	cacatgcaga	ggacccgagg	tcagaacaga	cctgccatgt
148141	ttgaggaaa	ggaagaaggc	caatgtgagt	ggagtgtgga	gagcaaagg	tgggcggggg
148201	cgagggacag	gagggcagag	tgggtggcct	gggtggggt	caggaggtca	gagtgggtgg
148261	cgggggcaga	ctgcacagag	cttctctgag	tggggggagg	agattggatt	ttctcctctg
148321	tgagtgggc	agccctgga	gggtgggaag	gacctgattc	ccctgtcagt	tgctctgggc
148381	atgttgtgag	aattgctggg	ggcaggatct	ccctccctct	ctctatgaaa	ttctgtttca
148441	gcagggatct	gcctatggct	gtatttctgc	tggtagaagt	ccctccctga	ttataaaccg
148501	ctcaacacat	agagaggtcc	ctaggagaag	gttcagggtc	cagtaggatg	gtgtgagcac
148561	gcctcacctg	catccccac	tgctgcagc	caccacaatg	cactggattc	atgacagct
148621	agctgaggac	cctgcaaaa	aaataacagc	atgcagattg	gggaagaaga	ctagaatgtg
148681	aggttcagca	aaccgcgaga	gtttactctt	cattctctgg	tattcctcag	cctagataca
148741	aggcaggtca	aactccagaa	gcgggcactg	ggtacagaca	gacagagctc	gagacaagcc
148801	ctgttgttgt	agctcaagga	tagaaaaagc	atttccgaat	gctcagagcc	aagccaacac
148861	ctcatttttc	tttctatttt	cctttttctt	ttcttcgtgt	tctcacacgc	cagctcatgg
148921	cagccacag	gcacctaaaa	ctctaagagg	ggagcctggc	tgtgtagctg	gaagtgtctg
148981	ggctcctaga	agggtggagt	aaacccatt	gctctttcct	cttttggctc	tcccaggagt
149041	gtagtgtcag	gaagtgtgca	agacagcagg	gtcaataaag	ccccagctt	ctggccagag
149101	agatttaaca	aggagccgca	ggaactgga	cagggttagg	aagatggcag	agacagcaga
149161	ccttggggaa	ctggcatcag	aaagtgtggt	agaaaactcct	gggctcgtgc	ttgagccgtg
149221	cacgagtggt	tctgacccta	aacaacacgc	cataggttct	ggggactgac	ttaggaggtg
149281	gaccaatgcc	caggtcccag	acgtgccact	gggtgtgtgt	tacatgggat	tgatcagaat
149341	aggactgcac	aggctttgaa	aatggacctg	acattgaacc	gcaatccaca	ggcacattgg
149401	aacttgcagc	cgaaccccaa	ttaacccaag	gggttaattac	ctgatcaaac	aaaaataatg
149461	accttctcca	tagcatttaa	acaagatcca	aagtctcata	gggtaatatc	aaagtgtcca
149521	ggatacagtc	caaaattact	tggcatacca	agaaccagga	aaatctcaac	tcctccaggc
149581	aaatacaacc	aacagataac	aagactgaga	tagcacatat	ggttgaatta	actgacaatg
149641	atcttaaagc	agctattata	aaaatactcc	ataaagtaag	ggcaaaaact	ctggcaacta
149701	atggaaagat	agtctcagca	agaaacggaa	gacataatgt	taaaaaggac	caaacggaaa
149761	ttttagaact	gaaaaataca	ataacagaaa	ctttaagaac	tcactggata	ggctcaagag
149821	cagaatggat	atgacagagg	aatgagtcag	tgaacttgag	gatagagcaa	tagaagttac
149881	ccaattcaaa	ctgcagcgag	taaaaatatt	ttataatgcc	aagagcctta	gagactgtgg
149941	gaaaataccg	aaagatctaa	ctaacagtca	tgtcccagaa	ggagaagaga	aagaatgtat
150001	tactaaacaa	atatttggat	aaataattgg	tgaaaacttc	tcaaatctct	tgaattgc
150061	aaacctacca	attcaacaga	tcggcaaac	tcaaatgaga	gaaatccaag	gaaacacaat
150121	cccacacgtt	ataaacacac	tctgaaaact	atagacaaaa	aaaagtcttg	aaagcagtca
150181	aagaaaaaca	gtgcattact	tataggggaa	aaatgggtca	gataactgca	gatttcttat
150241	caaaactgtg	aagtccaaag	gaagtgcata	acattgttaa	aatgctgaaa	aaaaaaaaaa
150301	agaactgtca	accagaatt	ctatctccag	caaaaatata	cttcaagaat	taaggagaaa
150361	taaagacatt	ctcagatgaa	gcaaaaactta	tagaatgtgt	ggccagcaga	tctaccctaa
150421	gatgggtgct	gaagggaagt	cttcagacag	ggggaaaatg	atgctagaaa	gaaattttga
150481	acatcaggaa	taaagggaag	gcaacagaaa	tacctgaaaa	aatataacag	attaggctct
150541	tgcgatattt	aaagaggaat	gataatgctg	tctgatgtag	ttttccatgt	atgtaggtgt



150601	ataaaagaca	acaacattca	ggggaggaca	gagggaccta	tatgatggca	gggtctcaac
150661	attcaacttg	aagtggtaaa	atattgattt	gaaaaagact	ctgaagagtt	acataggttt
150721	attataatcc	ctggagcaac	cagtaaacac	tatacaaaga	tatacaacca	aacacaatag
150781	gtaaatataa	attgaatacc	aaaacatgct	caaatagaaa	aagcaaaaata	agcccaaagc
150841	aagtagaagg	gaggaaatga	taagagtaaa	aatcaatgaa	actgaaaata	tgaaaataat
150901	agagaagatc	aatgaaacca	aatttggttc	tttgtaaaga	tcagtcaact	gataaatctc
150961	tagtgagact	gaagaagaga	aacagaaaaa	aatatgaatt	accagtatca	gaatattttt
151021	aaaaaggata	tcactaatga	catcacagcc	ataaaaaaga	taatggaata	gtataaacia
151081	ttctatgcat	ataaatccat	cagtgttgaa	gaaatgcatg	atctcctcaa	aaactataag
151141	ctacccaaac	ttacccaaga	caaagtagat	agcataaaaa	gttctgtaac	tattaaagaa
151201	aaagaattca	tagtttaaaa	tcctgcaaaa	aagaaatttc	caggtctaga	cggttttact
151261	ggaaaattct	actaacctct	caggaaaaaa	tggcaccaat	tctcccattc	attttatgag
151321	gtcaacatta	ccctgatatt	aaaaccagac	aaagacagtt	caataaaaaa	aaaaacccta
151381	caaaccaata	tcctcatga	acctataaac	caatatccct	caatgaacat	agacacaaaa
151441	atcctcagca	aaacactggc	aagtctagtc	cagccatgta	taaacagcag	agtgcacaac
151501	agacaagtca	tgtttgctct	ggatttgcag	agtctggcac	aatatccaaa	agtcagtgtg
151561	accagcata	ttaacagact	gaagaaggaa	aaaaaatatg	cttatatcaa	gtgatagaga
151621	aaaccatatg	atatgattca	ttcatgacta	aaaattaaaa	actctcaact	aacagagaaa
151681	gaaacgtgct	caacctggta	agaatatcta	caaaaaaact	aaagctaaca	caataccgca
151741	tcctgagaaa	ctgttttccc	ctaaaataag	gagcattttc	acgactccta	ctcaaccttg
151801	gggtgagtc	tagacagcac	ctctgtacac	accagcacc	cagcacagac	cctcaggcca
151861	ggcgaggaaa	gggccccaga	gagcttcccc	aaaaccctg	ttcctgggga	aggtggatgt
151921	ccaggaagg	tgactggccc	ctaaagccta	ggggaagcgg	gggcccctct	gctccagcct
151981	catgaagctt	tgatctccag	tgccaagccc	ggtgccagca	gagagccagg	cagatgaact
152041	caaccaggga	aaggacaccc	aggaaaggaa	caactccatg	agatcgagtc	tcatagcacc
152101	aagtgtctcc	acaggctctg	aggacatggc	tcttattcca	ggaccaggcc	tgacaggagc
152161	acagcctgag	ggatgagagg	gcaggcgagt	cttctgctgg	gtgggggaat	ggccaccagg
152221	gccagtccca	catctgttgt	ggccctggca	ggatcaaccc	tgggcctgag	tccttgagg
152281	tgaggccagc	acgtagtaga	ggggcagccg	agtaagaaga	atctctcggg	gaagtcagt
152341	gccagccagc	ctaagcccg	cccagaccct	cgccctgcta	gcatgagaca	gacacagggg
152401	gcccgcactg	cagagctgcc	ccggcccacg	cgccctgccc	ctgctgtgct	ccggccatctg
152461	ccccagcagt	gctgtgagag	catcgacccc	tgcaagcctc	agccttcagg	gacggcaaga
152521	cctggggccc	acaggaggctc	caccggccct	ctcagggctc	gagcaccggg	ccaaccctgg
152581	gtggccactg	ctcagcactc	caggacagct	ccatgacaca	gaagcagctc	tgcccatgag
152641	cattcaacag	cccggctccg	ggggcaccac	cacgcagggg	cttcccgggt	catgctgcac
152701	catccactaa	tagcaccatc	acgtatgcca	ggctgggagc	ccgggcgtcc	caagcctgtg
152761	acctcattct	ctatgtcct	cacggctgcc	ccacagaatc	tctgaaacac	gactcagacc
152821	tgtcactgcc	tgcccaaagc	cctccagcgg	cttcccacca	catgactgcc	ccgggtccccg
152881	ctgatttctc	cttcaactcaa	gtgccagccg	tggagatctc	ctctgtgtcc	ttggcacact
152941	tcaactcagc	acctgcctca	gggcctttgc	acctgctgct	gcctccacct	ggaatggcct
153001	tcttccagaa	gtccaggctc	tagctccgat	gcctcctcac	aaggaccttg	gtgaaggggtg
153061	tcccttctcc	atcacactcc	acactgccc	cctgtcctat	ttttttgtaa	caccatttcc
153121	caatcatcga	cacgtgggtt	ctttattggg	tcactctctg	ccccctgacc	agggatctta
153181	cgagagcaaa	gttgtccttg	ctgggtccac	gtcaccagac	cagtgcctgg	catctcacag
153241	gcatggagtt	gagtaacggga	ggtcgccggc	cctcccagag	ctcccctccc	ctacttcggg
153301	atggtgtaag	gtttccttgc	caaggtccct	gtggtgcaga	gtgggccaag	tttgccgctt
153361	tctggagcga	caacaaggag	aaaacagggg	tttaagtggg	ggatccgggtg	ctctgatttc
153421	tccccagtta	gtaaagcatt	caggctgctc	ggtggagggtg	ggaagacagg	cgctgagagt
153481	tgtctgacac	aattgtctta	gcagcaggag	acgtgatggg	gatgtcggga	gtcaccagga
153541	ccatctgaca	aatgaccgcc	ttggggatcc	gtggagctca	gcgggtgccc	gagagagcac
153601	tcgaaatgcc	agctggaaa	gtcacgcacg	ccagctggaa	ggggagccgc	gcagcccttg
153661	ggcaggctga	ggacaggaac	gggcagccgc	cactgttcac	gagcaagggt	gcaggggcac
153721	cagagcccac	aggagctcct	gcaatgcgat	gaccatgcct	gcaggccccc	ataatagaca
153781	gcctcatttt	ttttttctag	gcaaaaacttt	gggaggatcc	tttttaattg	tatttgcaaa
153841	acagtttctc	ccaacatgat	gcccattgtg	atgtggattg	ggggaaattc	ccaaggcctc
153901	tcagagaagt	gcctccttcc	taaggtgggtg	tccagcaga	tgaatctggc	tgggcgtcctg
153961	agtggaaagc	gggcgcgggc	tgccattggaa	ccccaccctc	ctgcagggtc	ctggcccttc
154021	tgctcctct	tccttcccac	cccaggggcc	ctgggatgca	gctaggccag	ctgggagcca
154081	gacagccact	gaagccctcg	gggtactgtt	ccctgccttt	agcatccccc	tcagctgcc
154141	attcattcag	ccattcttct	aggaggacag	gccaggccat	ggtgaggccc	tggttccagg
154201	gaaagatacg	aagagggtcg	aggcgggcgt	tcctgccttt	gggtctcttc	cgctgccctg
154261	gcagccccgt	gcctcctgac	tgtggactgg	agttctgcag	ctgcgccatc	cacaggaagc
154321	tacatggaga	gatggggcag	ctcagcagag	aggcgcaaac	accaattcag	ttgaaaactc
154381	tgtgttcccc	gggtagaggc	cagttaggag	ccagcatgcc	tcccaagggt	tgtgcaggga



154441	gacttcagcc	aggcaccac	tggggcagac	ccagtcccac	tgcttctgcc	ttcagtggcc
154501	acaggcctgg	ggccattggt	ggatagagct	ggagaaaggc	cacctcttcg	ggacccaggt
154561	cacaggatgt	ccccccaaag	ccatcccagt	cagatgtctg	cagacacctg	ctcccctggc
154621	aacagtgacc	cagacctgac	cttcaactct	aactgtgcac	agcttcctag	cccgcccccc
154681	aagcctgaag	cctggtccct	ccagacactg	cagaaagctg	gggctgactg	taatcccagc
154741	actttgggag	gccaagggtg	ggcgggtcac	gaggtcagga	aatagagacc	atcgtggcta
154801	acatggtgaa	accctgtctg	tactaaaaat	acaaaaaaaa	tagccacgca	tggtggcacg
154861	caccggtaac	cccagctact	caggaggcta	aggcaggaga	atcgcttgaa	cctggggaggc
154921	ggagttgcag	tgagccgaga	tcatgccatt	gtactccagc	ctgggcaaca	cagcaagact
154981	cagtctccaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaggctc	tggccctgcc	acttgaccgg
155041	gtgggtcacc	tctccgtgcc	tcagtttctt	ctatttaaaa	tgcatagggg	aggagctctg
155101	agggcaaaa	accctgggtt	tgccctcagg	cagggcagga	ccctggacc	agcaggcccc
155161	cactgcatgt	ccctcctgtc	ctccctgccc	cacccctgcc	ctctggcttc	tggggttcaa
155221	gtcagggagg	ccttgggagt	agcctgcggc	cctggggaca	gacaagggtg	cccctgaatg
155281	tcctgtggat	ctgcccaggt	tgatgccttg	actgctcctt	cctgtcacc	aggaccctcc
155341	aggagagacc	accccgact	acgcctgccc	agggtccctt	gcacgtccca	tgccccaca
155401	gcgctcaggc	gctgagatgc	tctgctgtct	gctgtggggg	cctgtggacg	gccctcctcc
155461	cccaccaggc	tgacgtccc	agcgggcagg	gatctaggca	tgcttgcctc	ataccacaag
155521	gcagaaacac	agtgaagaaa	gacagagaa	tggtcttggt	gtctttgggg	gctcagcccc
155581	ccagctccaa	gctggaggat	caggaggcat	ctacaagcct	ttccgaagcc	ccccatagg
155641	ctgccttgca	gggtacagta	tggcctagag	agccccagag	ggctcccctg	gtcatataat
155701	atctcaggat	ttccaggaac	aaggagaggt	tcaagtccac	agaagaatct	gattggcgtc
155761	tctggggcac	ccaccggtgc	tagaatcccg	gctgggggtg	gggatggtgg	tgacagacct
155821	ggctatgcaa	aatgtggcct	ggagcagcct	gggggcccc	gggagcttgt	tagagatgct
155881	gaatccaggc	cccattccaga	ccagctgaac	taaaacctgc	atthttggca	gttcccaggt
155941	ggtctgatga	cattgaagtc	tgggacacac	tgctctagac	cctcccagg	tcctccaaag
156001	gtgggtgtag	agggcctact	gcctgcctct	ggggacgcag	aggcatcagg	gccttagtcc
156061	tcctggggac	agtgaagggt	ccaccacccc	agaccatggc	caaactgcag	gggtcaggga
156121	ggcatggccc	aggacagcag	ctcagcagg	tgcccaaccc	cacctgcccc	tgctgcccc
156181	aacccacttg	ccggctgggt	cctctggacc	tccgcccctt	tccaccccag	gttggagcca
156241	aagaccaccc	tctccactcc	ccagcccccc	accccccaac	ccaccgagt	caggcccgct
156301	tgactcctat	ctcaacaaca	acaggaaacc	atthttccag	acagcattat	ataaggggag
156361	gatgcgtatg	ttcagcagat	gttactcttg	aggccgagcc	tccccgccc	tctgaacagg
156421	aaggccacgt	ccccatccct	ctcccccagg	gcctggcgct	cctggcccag	gtggagctgg
156481	cctctggctg	gccacttctc	tggaaaggcc	tcgtccagct	ctgtcctgtg	cagggtcatg
156541	ccctccact	ctgctctctc	tagccccagg	gcaggcgccg	cctggctcag	aggcacagca
156601	atggggccagg	gtcagacagt	ggtgagcagc	atgccaagaa	tcacagggac	atccacctgc
156661	cagcaaagcc	tcaatgacaa	caaaggacct	gaacagtggg	cactgacagg	gactctcctg
156721	cttgaccaaaa	cttttagtcca	gctcctgaac	cttctccttc	gcccacttcc	ttgtaagatc
156781	cagtttatca	agaactctgc	taagtacaga	atgccccatc	tctccatate	tgatcacctc
156841	tgatattctg	tcaggctcgt	cctccccatc	accctctggt	tgatgtctgg	tcaccccagc
156901	ctgtcttcag	ctagaatcat	gttaggtgag	tttagctaga	atccccgacc	cctgatgttg
156961	ctcttagtaa	ttccccatcc	cctggcccc	accctgctcc	ttggctatca	ccccccactc
157021	gcccgtgctg	tattcgaggt	tgagcccagt	ctctctcccc	gactgcaaga	gcccactcta
157081	gcggttccctg	tgtctatctc	catggtcctg	aataaaagtc	gccttaccgc	gctttaataa
157141	gtatcattga	atcatttttc	atttaacaac	actctgcctt	gctgaaggga	aggtagaaat
157201	gacagcagtg	gggaggagg	agaaggggca	tgggaaccct	gtgccgctaa	gggcatactc
157261	tagaccctct	gccccaggct	aatgttagaa	gctgcagagg	tgcaaatata	acagtacagc
157321	gaccatggca	gtggctcagct	gaaaaatggc	ccccgaagac	acgtccgggt	cccaatccct
157381	ggaacccata	aatgtttacct	tccatggcaa	aaaagaactt	tgtagatgcg	gtgagtccaa
157441	ggatcctgag	atggagtgc	tgctggattt	tccgggtggg	cccagggata	gacagaatgt
157501	ttatatcccc	cccaaattcc	gatgttgaaa	gctactcccc	caagggtgata	tattaggaag
157561	tggggccttt	gggagggtgat	tatgtcacca	gattccatcc	acacgaatgg	gatttgtgcc
157621	ccgatcagag	ggacccagga	gagctccctc	accctttcag	catgtgagga	tgcaacaaga
157681	tgccctcttg	aaccaggag	cagaccctca	ccacaacat	atgcacagct	tccagaaccg
157741	tgaataataa	gcttccgatg	cttttaagcc	gctccgttta	tggtatttgg	ctatagcagc
157801	ctgcatggac	taagacaggc	cctaaatgca	atcacacttc	tctttatcag	aggaaggcag
157861	agcaagatga	gccaggcaga	ggagaaggcc	acgtacaggc	agagcagagg	gagatggagg
157921	atgtcggccc	tgaagcccag	agcgaagagg	ccacaagcca	agcactgctg	gctgccacga
157981	ggacctggaa	gagcccagga	cgcactcctc	cctagagggt	ccggaggagg	cgggcccctg
158041	ccgccacctg	cgttttgccc	tgttgaaact	gatttcagcc	tccagaaccg	tgagaggata
158101	caatttcta	gtttcaaacc	agcaagcttg	tagtaatttg	caggcagcct	cagaaaccaa
158161	tacaatgggg	aactctgata	tggactaaaa	gtggtgagag	aattatggta	gaagaatggg
158221	gctgggagg	gcagcagtg	tgtcagtgag	ctaaaccctc	ctcttccagg	gcagggaagt

158281	aatggacaac	atcgaaaccc	gatacatcaa	gaaacagcaa	tagaatagca	tttcaggact
158341	tcatggcaac	cccagaacca	gccagcagcc	taaagcctta	ccggggctcc	ctcctcagga
158401	gtggcctgtg	aagaggggac	gcacacgcac	tcacactcat	gcacacacac	acaccgagac
158461	agaggagggc	ctggccagag	caggggtgtc	agcagagggg	acgcacacgc	attcacactc
158521	atgcacacac	tcacaccgag	acagaggagg	gcctggccag	agcaggggtg	tcagcagagg
158581	ggacgcacac	gcattcacac	tcatgcacac	acacactcac	atcgagacag	aggagggcct
158641	ggccagagca	gggggtgtcag	cagaggggac	gcacacgcac	tcacactcat	gcacacacac
158701	actcacaccg	agacagagga	gggcctggcc	ggagcagggg	tgtccgcaga	ggggacgcac
158761	acgcattcac	actcgtgcac	acacacactc	acaccgagac	agaggagggc	ctggccggag
158821	caggggtgtc	agcagagggg	acgcacacgc	attcacactc	atgcacacac	acactcacac
158881	cgagacagag	gagggcctgg	ccggagcagc	ggtgtcagca	gaggggacgc	acacgcattc
158941	acactcgtgc	acacacacac	tcacaccgag	acagaggagg	gcctggcccg	agcaggggtg
159001	tcagcagagg	ggacgcacac	gcattcacac	tcgtgcacac	acacactcac	accgagacag
159061	aggagggcct	ggccggagca	gggggtgtcag	ctccaggacc	aggagtctcc	aggcagaacc
159121	tgatgggtgg	gcgggggggt	cctcacacac	cccacacccc	tgcctatgcc	cagacctcca
159181	gcttcgggct	ctgcatccac	acggaggctt	cggattgggc	gcagcctggg	agcacacaga
159241	ccggccaggg	ccagggccag	gactgtgact	tcccgggctc	ctgggtcccat	cctgccccct
159301	ccaccgggtc	ggccctgttc	ttgtcttgtc	cagctcgggg	gaagcacctc	tcgccagcca
159361	tctggggctt	ggctgggcct	cagcctcctg	gctgccacgt	ctcactggga	gctgggcggc
159421	cgccagctct	gagctaagtg	agcgcagagg	agagtgcagg	gagcggcaag	gccaggctcc
159481	tcccacccc	gtggcgtggc	gtggcgtggc	cggaatactc	agttttccct	gctttttgag
159541	acaagagtag	taaattgcacc	gtgctgtact	cagagccacg	gaagcagccc	ccccaccac
159601	agcgtctcct	cctctgcagc	cccccgggcc	gtgtctctct	ctcacgggcc	cacatctctc
159661	ctggactgtc	tcccttttac	agatgaggtg	agacagcatg	cagaggccca	tgtgcttgtc
159721	ccacaggggt	acggggccaga	attacaagcc	cctgattctg	ggggcttcca	cgtctttcca
159781	ccaagccctt	gtagccttcc	tgcacccagc	tggggctcct	gtcccccgcc	agtgccccca
159841	acctctactc	cagcagacct	ccacctacag	acaggagtct	tgggcacaca	cttttttgtg
159901	ccaggcccaa	ccaagagctg	caagggacat	aggaatgtcc	cttagcatgg	cccgggagcc
159961	tcaagcccag	ggcctaccta	gaagtaactc	tctttcctga	cattaagcga	ctccccctcc
160021	agagccta	acagccagc	tacgccccct	tctcccagtt	attcaagcct	ggaaggaacc
160081	tggaggtggt	actggattcc	tacacccacc	ccccccacaa	tccccaaagg	cccaccccg
160141	cctcctcctc	ccccctgccc	tgcattggagc	aggcctcact	cactctctct	cttctgttcc
160201	tcttgggtcc	ctgcctcaga	ccctcccacc	taacacagcc	tctgccccac	cccagagggc
160261	tcttcccaag	acacgcgtcc	acgtcgccat	ggacaagatt	aaataccctg	gcttctggaa
160321	agcactgagg	ccacagttgg	cttctggaaa	gcactgaggc	cacagttggc	ttctggaaag
160381	cactgaggcc	acagttggct	tctggaaagc	actgaggcca	cagtgccttg	ggaaggtcat
160441	gtcgggatcc	cagggctgac	tgcactcctg	ccagtgaagt	gggggtgtcc	ccaaaggggg
160501	tttgtttccc	accattggga	gccaggatcc	caggtcagga	aacagagctt	gggtgggagg
160561	gaagccatct	ctggggactt	tcttcccatc	tcccttgctc	cgtctcacag	gagaagccag
160621	agctagaagc	tggcatgagc	cctcctagca	ccaccagcca	cagtctccac	ctccggggac
160681	acgagctggg	atctgagtgg	cgcaggctact	atgttctaga	gagggcgtac	ccaggaccgc
160741	ccatggcacg	cagaggaaaa	ctgaaacaac	gatccttcca	gagacaaagc	acagccccag
160801	atggccacag	gaactggccg	cccccaggag	ctgctgagca	caaacgccac	tggagagagg
160861	gcttgggtgg	tgcaccgaag	agcagaacag	aggaactgaa	cagccacagc	caccgcagaa
160921	ctgggggatc	tggacagagc	tctcgctgtg	tgtctaaaga	ggagcaagag	acagagggca
160981	gggagaaaa	ctgcgagacg	tgygcaagtc	ccagaagaga	gtcggggagg	tgagagaagg
161041	aatattttcaa	aaaacaaaat	tacatgaaag	attccagcat	ggatgggttc	ctagaggcct
161101	tatcaataca	tatataagaa	aaagctccat	agctagaccc	acgataatga	aacctaagaa
161161	catcaaagac	aaagagaaa	ttcaaaagct	cccagagggg	aagagcagat	cagatcagtg
161221	taaacaggga	ggcagcaggc	agacatccaa	gtcccagact	ggggcgtggg	caccgggcag
161281	cagacgggag	acaagagccc	aggaagtcct	gctggcaaa	aattcaaaag	ctcttgccct
161341	gagggaaag	gctgccccat	tccgaggcca	gagccaccgc	agcagaggac	agctgtgtcc
161401	ctcagagggg	tcggcaggga	cagaggccac	accagagcag	cagagatcta	gcaacccaga
161461	agggtgacac	cctgggagat	gctgagagca	gcagggccct	caggagcagg	gctggggagt
161521	cggccacagc	tcagggggtc	attgcattgt	ggcatcccag	gctcctgcct	aggcccccta
161581	gggtctccca	gccccacaga	acacctccc	cactcccagg	gatggcattt	gagaccttcc
161641	tgtgtgttcc	ctccaggaac	ctgtgagctc	caagagaaca	gaaggctgcc	tcttcaaccc
161701	cagtccagca	aggccagggg	cagagcaggc	acggagaact	gctgtcgaa	atgaagtggg
161761	gacgggacgc	agtggtgtac	acctataatt	ccagcacttt	gggagggcaa	gaaccgctta
161821	agtcacaga	ttcaagatca	gcctgggcaa	catagtgaga	ccccatctct	acaaaaaata
161881	aattttaaaa	attagccagg	catggtgggt	cacacctgta	ctttcagcta	ctttggaggc
161941	tgaggcagga	ggatcccatg	aaccaggagg	tcaagggttg	agggagctac	gattgcacca
162001	ctgccttcta	gccccgggag	gagcaagacc	ctgtctcaaa	aaaaggaaat	atattttaatt
162061	ttttaaaaa	aatacacagt	ggtgagaaa	gatggaaa	agagaggaaa	ggaaggaaa

162121	gatagggaaa	gaacaaaaga	aggaaagaaa	gaaaltccag	ctaattctca	gcaacagaca
162181	attacaggaa	cacaaggaca	aagcccagag	ccccaacccc	acgtggcctg	gggtcccaca
162241	gcctccagcc	ctcctgtcac	tctgccagct	ggcagcaggg	ccacctctga	gcattccttt
162301	cctcagcact	ttggcgtctt	tcctcccagt	ggtttcttgg	tagttcctca	taaatccatg
162361	cgtlaggtctg	gctacagcta	tccttgtccc	tgaagatgga	ctcagctagc	agagtgcgtg
162421	atatggtaac	acaatatcac	acatctcctc	caagcctgct	cggacgctgc	ccttctgacc
162481	ccgatttcctg	tcaccaaacac	aacaggaaaag	gagcttgaag	acccaggagg	ccacagcaca
162541	gttcctcctg	gagggagaga	gatgagagca	gggtggcgga	gggtgcactc	atgagacttc
162601	tataagcacg	gacccagagg	ccatcccagc	gacactgtgt	ggggactcct	gcacctcttc
162661	ctgctgcccc	gaccatgggt	ggccagcagc	actgcagtgt	tggtgtaatc	tatgtcccc
162721	ggatggccag	cgggcacaga	gcatagtgt	gctgagggac	acatgcccc	cacaaagcc
162781	cccacatgtg	catcgcatac	atcacccata	cacacacgtg	tacatttcag	cacgcgtcta
162841	ggggaggtta	gcgtgccagc	accagctgag	actccctggc	cagactcatc	gccagttggg
162901	gagagtggag	cagacgtaca	ggatggaggg	tagaggtggg	gctggtccac	agtgaggagg
162961	ggaggcaggt	gtagactggg	ccagggaag	gttctggaaa	cccagactga	agggtccctg
163021	gggtccctgg	ggctcatcca	gggtcgggtg	ggagccagga	catgaggaa	gggcagtgat
163081	gaatctgaag	ccctgctggt	cagcagagcc	tccacctgca	cgtcctcccc	aaaggctcag
163141	caggcttcac	cctggccacg	tctctaccct	ggccattagc	aaagaggccc	cccttgacc
163201	aggaggaaca	cagcaggcca	gcctcctgct	ccttcccata	caaggctcca	gccactctca
163261	tggagactct	gagctgagac	actcccagca	gaaagggtea	tttatctcag	agcttaacta
163321	gccccttcca	attagtggag	aataaaatgg	cagcatttac	tctgaagaac	tgcgtatcga
163381	gagtattaat	taaatgcgct	gccgctgagc	ctcagcacac	aggaggctga	gccctatta
163441	aaaaggcatg	agactaaagc	aagacaggga	ggagaggcgg	ggccagctag	acaggggatg
163501	gggagagaag	gagatgccgg	gcggggcggg	gacagccagg	gacggccaca	gctgccctgc
163561	caggagggac	agggccact	gcaaagtggg	acagctgctt	ctagccctct	cacaagaagg
163621	ctggtgggga	agaagccaa	cacacgcaga	gggtgctccc	gcacaccct	cgctccgtc
163681	tttttggaat	cttgacagtt	agaagacgac	aagggccact	gtccagtggc	tggcatttgt
163741	ccccatgaga	ccacagagag	gcagaggaca	gctcctcttc	agaccagggc	agggtgggca
163801	atgccaaagg	cccctcggt	caggatggag	ctgagcctcc	ccaggatgac	tcccctgggg
163861	gaaacactga	tgctgctgtt	tgagcgttga	gacgttgggg	cctggcctct	gagaggcggg
163921	gccccgaagg	gtcttccagg	aaggtcagac	cccagcccaa	aagcagcagg	ctggggccac
163981	aaagagagct	gggtccctggg	ccagaaagcc	ctcttccacc	atgcactctg	ggcaaggggc
164041	gtgtgtgcc	ttaggaacac	agctgcacgc	tatttctctc	ctgtcttctc	ccagctgaag
164101	atttcacaca	gccctgttat	gactgagggt	tcccaacagc	tattccctca	acacacaggc
164161	ccccacacac	atgcatgtgc	actcatgcac	tcccacatgc	agcatgcatg	cacccacata
164221	caaagtcca	tacgcatcac	atgcaccacg	tgtgactca	cacgcaggcg	tcacacaggc
164281	accatgcaga	cgcatgcaca	cacacatcac	tcatagtgct	atctcctgtg	acacatgcat
164341	cacatacaca	tgtccacaca	tgtgactac	acacaagtac	acacatgcac	cagaaaaatg
164401	ccccatgcat	gtgcatggca	gacattaccc	acacatacac	atgtgcatcg	gtcatacatg
164461	tgtatcatat	gcaggatata	tatgtacagg	catgcagtgc	acacatgcat	gcataagca
164521	catgcattat	acatgcacac	gcccgcacat	gtgcacacat	gcatagacat	catgcactca
164581	cacatgcctc	acatttgcac	acatgcacat	ctgacgtgct	cacacatcag	ccacacacat
164641	gcactcactc	acaccctcct	gctgctccca	gcctcactgc	ctgcttgcct	actttcccag
164701	ttcctgccac	tacctgagaa	gggtgtttat	tggccttttg	cttattgtat	gtctctacca
164761	cctcatccta	cccctcccag	ccctccccat	ccacactcaa	ccagtagaat	agaagcctct
164821	cccctctggt	gaggcaaaaa	gctgtgaccc	tcttttcacc	cagaaaccac	aggaagcatg
164881	cgtgcagact	ggactcagcc	ttcaccagcc	cccgcagtca	ttacgcctgg	ctcccagca
164941	cccaaaccac	gggtcccact	ggtcagcacg	ttcacaccca	gcctgccaca	tctccggcct
165001	ggcccacaga	atgcctgggt	ccagactcct	aatgtaggaa	gactgccctg	cctgatctca
165061	aggcttgctt	ctctgcattc	aggccactcc	tgctgcccag	cctgacggct	ccgggagcag
165121	ccggtgatca	gtcaacctcc	ccagccttct	gtttgccttg	cagagccagc	cgactcccag
165181	gagagctgta	gaggctgggt	agcagcctca	ctgtccctgc	cctgctgggt	gaggggaagg
165241	aggccacgga	ggaaattccc	tcccgtccca	tgtctgcctc	gtgtccttct	ctctcctgct
165301	gtgctcctgg	gcagcagctc	cctcctccag	ccttccagggt	ctcccaatcc	tggccccagc
165361	cacttccctg	ggaaccttaa	gcctgcctgg	agtcacacct	cctccatttt	tttgggtttt
165421	ttttgagacg	gagtctcgct	ctgtcactca	ggctggagtg	cagtggcacg	atcttgactc
165481	actgcgagct	ccacctcctg	ggttcacacc	atcttctctg	ctcagcctcc	tgagttagctg
165541	agactacagg	tgtccaccac	tacgcctggc	taattttttt	tttttttttt	tgtattttta
165601	gtagagacga	ggtttcacca	tttttagccag	gatggtctca	atctcctgac	cttgtgatcc
165661	gtccacctca	gcctcccaca	gtgctgggat	tacaggcatg	agccactgtt	tctggcccac
165721	ccttctcctt	ctgatagcac	agccccctgtg	tcactcagga	accagctctc	atccctcgct
165781	caccagtggg	caaggagggc	acatgaccca	gcctaggcca	ggtagaatac	gttgtcatct
165841	gactacatga	acatagctca	gggatgagga	cttgccacaa	ggtattccat	gaagttaa
165901	ccaggaacta	ttgctggaat	tggtggaaa	gaaaagctgt	ttcccccggt	ttgttgcctt

165961	gtgaggaggt	aatgcagcag	ctgcctgcac	ccatttttct	gcctcagaga	agagcaggcc
166021	tgagagtga	atcgacatag	aggagagcaa	agccaagggt	tggggatgtg	ggaaaaaaaa
166081	caactgaaaa	tttatttggg	cacctggatc	cagccacacc	tgaagctaga	tagttctaga
166141	tatttttctt	tttgcagatt	aaccaataag	tttttttttc	ttttttctga	agccactttt
166201	acttggaatt	ccaacacttg	aggaaataaca	cctgactaat	acactctctc	ttctatagct
166261	tcaaactctc	catctcccgg	atcctcatcc	tcggccctg	agcaaactca	agccacctaa
166321	aatcctctgt	gctcgaggca	gggccaacaca	cacatctaca	caagcgacaca	cgtaaacact
166381	gcactcacat	ctaagactgc	acaccctccg	cagctgatgt	ctgccccacc	acagcatgtg
166441	tgccatgcgc	accagcttcc	ctccactccc	ttgcctctca	ctggcccctc	acctggtcag
166501	tctagcttcc	accaactccc	cagcgacctc	gcaaagcccc	ctcttggtg	ctgggcctga
166561	tgccccctgc	catcccaatt	ccctgaaccc	cctcctcttg	aacctctctc	atccttttag
166621	aggtgggcag	cccgctcctc	aagaccccac	tgccctcctc	acttgctctg	acatgggtgg
166681	gggccaagga	gcacttcctt	ccccttctgc	ctccctcttc	ctttggtgcc	cgccaccca
166741	ccccaggatg	tggaatccac	agcctcccat	acagagcgct	cccaaccgga	attctacca
166801	gaccttgccc	ctgcacctag	gacctgcatt	tccagctgcc	agactccttc	caagagacag
166861	ccccagtgc	aaaccctctc	tccaagctac	ccttccctctc	taccacctat	ctgtgcaagt
166921	caaaagctgg	ggagtgcctt	tcctcagcct	ctccctcagc	ccccaggccc	acagacccaa
166981	gtgcctggag	atcccggcct	ctccatgttt	cctgagtaga	ttatgctgcc	tctattgaaa
167041	gatgttccgg	cagaacactg	gatgatccag	aaatatttcc	attatgtgtg	gcatgggcag
167101	aggcagatca	caaagggaag	gtacaatacg	acccttttgt	cattaagtat	gtttacacat
167161	atgtgcattg	acgtgggaga	cacacaaaaa	tgttgaaagc	taatggcact	gcagggattt
167221	ccattttctc	cttttggctt	atttctgcac	tctgttttat	atgaaggcac	atagacagct
167281	ttttaactta	aagacaggac	aaaaagttgc	cagccatcgt	gagtcacccc	aggactgccg
167341	gtgctgggaa	tgccaaaatc	tgatgggacc	cagcgagggt	ggggagagga	gcagaggatg
167401	caccacgggc	ccgtgccttc	tttgaatcag	tcctcacggg	gccatctcaa	ggtggagact
167461	ctgaaagtgc	agagaagtga	aataaacccg	ccatggtcac	acaggaggga	gtgcagggtc
167521	gcttctctcc	tccccacccc	aggtcccgac	tacaggctct	gatccacgcc	ctcccttgag
167581	gacaaaccct	gagccacgct	gggtccctcc	gcaggccctg	aaatagagct	tcattgccac
167641	atcagacacc	gggagacgcg	aggaggccgt	gagccaattc	cagatgactt	gtcaaggctg
167701	cgccagccac	agagtcaggg	aagcctctgg	ctgcctctgt	aatataagaa	ccaggactga
167761	agggccctct	gggagggcac	tagacgaggg	cctcggcagc	tccgcaggca	ggtggggccc
167821	agggcaggcg	ggacagaaac	ccaatgtggg	cgatgcaaac	tggacagaaa	gaggtcctc
167881	tgggctgcgg	ctccctcact	cggggcctga	cggtagcagg	gggagttaat	ctctcacaca
167941	caagctggcg	ccagaccctc	aactgcagggt	gtccctaaag	agaggtgcct	ggcaagtcct
168001	ctcctgggag	tggacgctct	cctcctgagc	acggtgtgga	agcaggggccg	tgcccagcat
168061	gggtgcgtga	ggactggatc	ctgcctgcag	acaggactgg	cgggaggcag	aacttgagct
168121	gcctaggact	gaggccactc	cagagccaag	actgacagtc	tccacaaaac	caaggagccc
168181	tcacctgtct	catggctggc	aggggctgca	caggaagtga	cagaggcctg	gtgccctcga
168241	gaggcaggca	ccgagccagc	cagtgtcagg	gagcaggctc	gcacatctga	tggaaaatcc
168301	atccgtggtc	ccccttctga	caggcctttg	gctgccttcc	cttctaataa	gctgcgaagg
168361	gaaatggcct	ggcagaagga	aagacaatgg	ggccccaggg	gcagcagtc	aggagggcct
168421	gggcccccca	ccctccagct	gtaggactcc	caagctaagg	acagagagaa	tactttgggg
168481	aggagaaaga	agcagacacc	aggcttcgtg	gcagggtgca	gcatggctga	gggcaggcag
168541	aacagccgcc	agcttgatct	ggacacctgg	tcctgccaca	ggccacagca	cttgccaggg
168601	gccgttggga	cctgtcactg	ccctgtggga	gctccctgga	gagttcttaa	cctgggctcc
168661	ataaactggg	aaaagagaac	acttcaactt	catctttact	aactgggaac	tgaaagctgc
168721	cattcccttc	aactctggag	agaggcagca	accccagact	taacagcacc	tgcgcccgtc
168781	agggcagaaa	tcacaggctt	cggcaactgt	cagcaccgct	gctgcagacc	agaccaaagt
168841	gtgtttacac	tcagcacaa	ttaaatata	gaggctgtaa	gatgcactcc	taggtcttat
168901	tatttaattgc	atcaattaaa	agcacacaaa	aaagttttac	atgaatgttc	ctggcagcat
168961	tattcacaa	agctaaagg	tggaaacaac	ccaatccatc	aacccatgag	tggattttaa
169021	aaactgtggt	acatccatgc	aatggaatca	tatctttcca	tccaaagtaa	tgaagctctg
169081	atgcatgctg	caacctagat	gaaccttgaa	aacattatgc	tgagtgaag	aagccagaca
169141	caaagggcaa	atattgtgtg	attcatttat	actaatgaaa	tgcccagaac	aggcaaatct
169201	atacaggcag	aaggtaggct	agtggttgcc	taggcctggg	aggacgggg	aattgggggt
169261	tactaaagg	tctggggttt	cttttcagag	tcttgaacat	gatctaaaat	tgtggtgatg
169321	gttgaccat	tctgtgaaca	tactaaagcc	attcattgaa	ttttacgtta	taagttggta
169381	aaatacatag	tatgtgaatt	atgtctcaat	aacattatta	tcaaaaatgt	gtaaggtata
169441	ataatgatta	aaaattttta	aagcatgtgt	atgttgatag	ctttattttt	ctcttttttt
169501	ttttttttga	gacagagtct	cactctatag	ccccaaactg	agtgcagtgg	tgcgatctca
169561	gctcactgca	acctccacct	ccaggggtca	aacaattctc	gtgcctcagc	ctcctgaata
169621	gttggaacta	cagatgcccc	ccaccacgcc	cggctagtgc	cttgtatttt	agtagaggca
169681	gggtttcacc	atgttgccca	gggtggtctc	aaactcctga	gctcaggcca	tcctcctgcc
169741	ttggcctccc	aaagtgcctg	gatttatagg	gtgaggcacc	atgcctggcc	gatagcttta

169801	tttcaatcag	atttattttcc	tttgtattttt	gtagtgtaca	ctaaacacat	tatttttgatc
169861	aggggtccat	gggcctcacc	aggctggcag	gagggttcac	gacatggaac	agctgtgcta
169921	actaccagga	cagaagacag	gaaggtttgc	tggaagaggg	gtgatgagaa	gttggccagg
169981	cctatgggggt	tcccagcatt	caggcaaaag	cccagagggtc	ctcagggacc	ccccactgt
170041	ctgctctgct	aatgcaaagc	ccatgaggcc	caaggagtct	cagcttcttc	agccacccga
170101	tcacacccc	acacacacta	ggccctgggt	ccttcaagtc	cttaggtcat	tttcagattc
170161	aacttcagcc	aagacaaatg	gtcccctcct	gcacctgcca	caggcactgt	ttccaacgtg
170221	ggcagagttg	acagttgatg	ggtgtgtgtg	atctttcggg	gctcctgtcc	caggccca
170281	ggatgaacgg	gaaggtggac	atggtgggtg	gagcaagcag	gaaatgatga	tggtcaggag
170341	gggccttgaa	tgtctgttct	gggggctgt	gctttaccct	caaagtggag	gagccatggg
170401	gagcatttcg	ggtggggcac	ggaagggaga	cgcgtgctt	atcatgggca	gctggaaggg
170461	agtgaggaga	acgggggcag	ggagaaagaca	gggcagatgt	cctggagaca	gaaggcaatg
170521	gggacccata	gctcctcctg	acacatcagc	ctggctttcc	tctggggccc	atcctctccc
170581	cgacccactc	ccagagggtc	agaggacaga	cgccagccct	ggagccacag	cagtctgtg
170641	acgaggcctg	gccagccgaa	gtcctgtctc	gctgtcctgg	ccatagtgat	ttgcctgaag
170701	catgtgaccc	aacctgagcc	atgagaact	caccactggg	ctctcccaga	aacctcggg
170761	gaacctgggc	ctctctctgc	cggcgctgtg	gtgcaactgc	caggcaggac	ccacgggag
170821	ggcggttgag	ctgaaaccgt	gcgggaggcc	tgggtgccc	ccctgagccc	cgcattcccc
170881	ctttggcccc	agccgcttct	gccccggtt	tctgtcactt	gcagtacacag	gagtgtggac
170941	ctggggatgg	cctgtgggag	ggacaggggc	acaggcacca	gctgcacagg	tgtggatggg
171001	aggacaggag	cgaggccagc	cgcctctggg	ccacaccctc	cggattcact	cccaaaggag
171061	aacggccagg	cgcccaagtc	ctcccactg	aggggtgcgg	aggaagggat	cagagcgctc
171121	ccccaaacaa	ggctgctctc	ccgtcctgca	catccctcgt	ccacactgac	ctccatttcc
171181	cttcttgggg	cctgcccgc	acgagactct	tctcccattc	cagactggaa	atcttttccg
171241	caccttgact	aattgttttc	aggtgggaag	gatttgggtg	agaaatggcg	tttgtgtgtg
171301	cgtgtgtgtg	tgtgtgtgtg	tgtgtgtgtc	caccccagc	cgggcagctg	ccccaggaat
171361	gctgtatttg	cgggaagagc	tagaggaccc	ttggcctcag	ccatgtcctg	aggggctctc
171421	ggtgatgtca	gcggctcctg	gcctcttcca	gtaaacaatc	gttcttatte	tgaccagga
171481	attccatttc	agaacagtaa	cggcctccac	tggcaacaca	tcccaggccg	gtgatgcctg
171541	gggcaggagg	agcatectac	cccggcctgt	gccccgcccc	cggccccgcc	ccccactccc
171601	atgcccgc	aagccccgc	cacagagctc	tcaaaaggga	caaaatctgc	acctggctcc
171661	gtgtgtccaa	gcgctttcac	actgtgttcc	ctcctagcag	cgccatcagg	aaaatatgat
171721	cttgtagtaa	caggcccttc	gctgtcagtc	gctggccctc	gcttcattta	atcctcatca
171781	taaccaccca	gagtagttac	cgtcatcagc	ttttaccttt	cacagatgag	aaacaggagg
171841	ctcgctcgcc	cagggaggga	gggagcagaa	aaccaagccc	acgtgttggg	gcgagatcgc
171901	gtctgactcc	agcgacggg	accttgcagg	ggctgatggc	caaaaactga	agtcctggga
171961	aggagaggga	gagggagagg	gagagagagg	ggaggaaagaa	atgaccgttc	caaagcccat
172021	gcagcactca	tctcaaaata	agagaaactc	tcaggtgcct	acacagggag	tgaatgcgcg
172081	tgaagcccgc	cagggtaatg	atagggagca	gcagagactc	ccaaggcacc	tggaaactgtg
172141	ggacccttct	agagggttcc	atccttgtct	agctctagcc	ctccagcaca	tagggcccggc
172201	agatttctctg	gttctcaaga	gaaactggaa	atcttgagct	ttgtgagaaa	ttttcaaaaca
172261	ctggaaaagc	caaacaaaac	tcattctgtg	gtggacatgg	cccgtggccc	accagtgtgc
172321	aaactccgcc	ctgcacaatg	ggaacctga	agtgcagaaa	attatctcat	tgtgaagggt
172381	aaggcaggga	ctcaaaacag	ggctcctcag	tccacattca	tagagtctcc	aggttgtacc
172441	accgaatcta	ctcgccagcc	ttctcctgaa	agtgtctcac	tgaaccacat	cccccgtaga
172501	aactcagggc	atctttcaaa	gaccgcagtt	ttcctgcccag	cctgcaaaaca	acctgtgtca
172561	tgcccggtgc	actgctcccc	aaccacaggg	gacacctaac	agaccacgag	ggactgagag
172621	agcctccttg	actgaaagg	aaggtcaacc	gccgccactc	atccccataa	aaccacaaca
172681	tgggcacatc	ctgcgttcga	agctctactt	gatgcaagag	agcagggaaa	ggaaagccgt
172741	gccaccgtgg	ggaaatgggc	ctgggagatc	cattcaggcc	aaggtttgtc	cactttgggt
172801	ctaccaccac	cagcattcac	agaagaagca	gcacctgggt	ctcaattccc	ttagaagcgg
172861	ggctagatgc	tctgaaggca	tggttgtagc	tcccacacaa	aattactaaa	gacataaaac
172921	gaaaacatgt	tgaaaaagaa	gagacaatat	tatgaactgc	agaaagtaac	tgcacacgta
172981	gagaatcctc	acccccaaaa	acatccaaaa	aaaggaaacca	aataataaca	aggaagtcca
173041	taaaaagtct	agatgcaaaa	taaatagaga	aactcatata	tctgggaacg	caaacatggg
173101	aaacatccaa	gcgagagatt	taggtttgtt	ttgttttgtt	ttttgagaca	gagctctcgt
173161	ctgtcgccca	ggctggagtg	cagtgggtgca	agcgatctga	gctcactgca	ggctccgcct
173221	ccggggttca	cgccattctc	ctgcctcagc	ctcccagagta	gctgggacta	caggcgcgca
173281	ccaccacgcc	cggctaattt	ttgtattttt	tagtagagac	ggggtttcac	cgtgttagcc
173341	aggatggtct	cgatttctct	acctcgtgat	ccaccgcctc	cggcctccca	aagtgtctggg
173401	attgcaggcg	tgcgccaccg	cacccgcgaa	gagatttatt	tcagaaggac	ttttgtgat
173461	tatctgtggt	tcatatattgt	gccattctag	taacagagag	actgctttgc	taagacatgg
173521	gaagatgcag	gtcattggga	gcaaaggatc	tgaagcaggg	gctgttgcca	ggaaggcg
173581	gccttgagga	cagcaatgct	ggggcctggg	cttgccagggt	caaccatggg	tatcaagtg

173641	ttcgtggcag	ctctgcacat	gcatgtatga	tttgactaa	tgtgtcattt	ttactctgt
173701	tttctccctg	ggaataaaga	tgtctaaatt	ctgagaatat	cattagtctg	aaagaagcag
173761	gtcccaggtg	ttctgagtaa	ataatcacct	actgtattat	aaattaaaag	tggtcccaga
173821	taacaatctg	gtgtaaaata	taatggaaaa	gaccctattg	acaatggcaa	taacattcta
173881	agtctcttag	taataaacct	actgggctat	gttttttaaac	atctgaatca	atcttcttat
173941	ttccgtgaga	aaattaaaca	tcataagat	gtcgttcatt	tggtcagcaa	agatttgctg
174001	agggccttct	gcaaagggga	cactacttta	ggcatttttt	taaaacctca	cgatggtgat
174061	acttgcccta	tcagatgtca	gaacatattg	tcaaccaaca	gcaattaaag	acccctggac
174121	tggtgcaaaa	gaaaaacaga	aaaggccagg	cacagtggct	catgcctgca	atcccagcac
174181	tttgggaggc	caaggagggc	ggatcacgag	gtcaggagat	ggagaccatc	ctggctaaca
174241	cggtgaaacc	cgtctctac	taaaaataga	aaaaattagc	cgggtgtggt	ggcgggcgcc
174301	tgtagtccca	gctactcttg	aggctgaagc	aggagaatcg	cttgaaccca	ggaggcagag
174361	cttgacgtga	gccgagatag	tgccactgcg	ctccagcctg	ggcaacagag	tgagactccg
174421	tctcaaaaaa	aaaaaaaaaa	acaaaacaaa	acccagaaag	gccagaaata	aaccccatat
174481	cctccatgaa	cttcacagat	gatggaatat	acattccaaa	gtcaagaaga	aagagtgaat
174541	tagcaaacga	atggtgttga	gacaacagca	tggaatttta	gaaaaaaatg	aacacttaaa
174601	accaacacct	tatacagtat	aagaaacaca	tttacttga	attaaaggag	ttacatat
174661	aaaaatcaag	tcaggaaaaa	agcaaaacat	ataatggagt	atttctttga	aaggctgata
174721	cttttctaca	ggctgagtaa	ccagaagaaa	gcaactagag	aggctgacag	agctgaccga
174781	agaaacaaaa	caccagcccg	gcagaataaa	aaaggaaagt	ctgcaaatat	aaaccacccc
174841	aagctgcag	aggggttaat	actagcccca	gggagaacca	atcagcttca	gaggaggaaa
174901	gttgggggccc	gtaaccacca	gccattaaag	aaacaagt	caggccgggc	gtggtggctc
174961	acgctgttaa	tcccagcact	ttgggaggcc	gaggcgggtg	gatcatgagg	tcaggagatc
175021	gagaccatcc	tggtctaaaca	ggtgaaaccc	cgtctctact	aaaaatacaa	aaaattagcc
175081	gggcgcgggtg	gcgggcgcct	gtagtcccg	ctactcgga	ggctgaggca	ggagaatggc
175141	gtgaaccggg	gaagcggagc	ttgcagttag	gcaactgag	gccactgag	tccgagctcc
175201	ggcctgggcg	acagagcgag	actcgtctc	aaaaaaaaaa	aaaaaaaaaa	aggaacaag
175261	tttcaacaat	aacacgattg	cctgactccc	accaagggtc	ccaggccagc	ccagccccag
175321	cggactcctg	caggccttgg	ggagccagg	gggcacctgc	aggcagggtg	ggggaaggct
175381	gtgaggactc	tgagccccca	gagtctatca	ggacaggctg	acaccaagac	agagacggcc
175441	tacatctacc	cacccatac	catgccag	cctggaccag	ccatccctgg	ccccaaaggc
175501	cacagcgtag	tcaccacagc	tgagactgcc	ccaatcagg	ccacctgcaa	tgtccagcag
175561	cctccctgca	actgccccg	ctttggtttt	taatgtccct	agtcccaga	atgcagaagc
175621	tctgcaaaaga	gaagaacgca	gcaactctgg	caacctgact	gtggaggcag	acacctggca
175681	cacatttccc	ctccccag	gcccgatgc	tccaccagc	gctgggcacc	ctccaacctc
175741	cactccattc	actcccagac	ggagcaaaag	ctggcctcat	tggccacctt	gcccagcaca
175801	ccacctgcac	aaccttccca	gcagccctgc	tcctcggg	accctcccca	gtgaggacac
175861	agcccgtagc	agcctgggca	ccccaacagg	aaagacagcc	caaaggccaa	gccgccttat
175921	ctgagggtga	ggatgctctg	gcagggcagg	gactcagcct	tgacctgtt	tgctagaaaa
175981	tgaccaaga	actcagtagc	aggccaagga	gaggagacat	gaattccctt	tttcagcaat
176041	tcaccctgct	gttacaggtg	ggaggttata	tgtacctctg	aggaggtaca	gatttcatct
176101	ggtgctccag	ctgcagtga	aggagctggc	tctgggctct	gtgggacctg	gggtgggtca
176161	tcccacatgc	tgtgccctga	tttctcccca	gcacattggg	atggctcgtt	gttggttccc
176221	agaatgtacc	aggaaggtac	aggtgcagag	tgacgtgac	gacgaagggc	ccccagccag
176281	gaagggctcc	caggagtcct	ggctgatgaa	agactctggg	aaagaaaggt	gggaacaggg
176341	agaggacaga	aggggcaggg	aggggagagg	acataaaagg	agaatggaga	gaaaagggaa
176401	aaaggggaga	tgagggagaa	gagaagggag	gggagtgga	aaaagaggaa	tgagaaaaa
176461	aggggggaaga	gagaggagg	aaggggagga	gaagctggag	ggagaggagg	gggaagggag
176521	ggagaggagg	gggaagggag	ggagaggagg	gggaagggag	ggagaggagg	gggaagggag
176581	ggagaggagg	gggaagggag	ggagaggagg	gggaagggag	ggagaggagg	gggaagggag
176641	ggagaggagg	gggaagggag	ggagaggagg	gggaagggag	ggagaggagg	gggaagggag
176701	ggagaggagg	gggaagggag	ggagaggagg	gggaagggag	ggaggggctc	ccagcaccag
176761	gaccagcac	aaccgtccta	tgacagttgt	ctctgcagct	gcatttaggg	ctgttctgga
176821	aactcaggcc	ttctccctgt	cggcagggct	gagccctggt	ctgggacctc	ccctgttct
176881	atcacagccc	ccaggctggg	tgaggtccac	tggaatctgt	tttactcac	aatgcttctg
176941	acacaaaaca	tgtggtttcc	ctcacacaaa	ccagttctcc	aacctccaga	cgctagctgg
177001	gtgccccaca	gttgagtga	atgctcacgc	caactacctg	gggtcagctt	agccctcagg
177061	taaaggcctc	agtccacac	agcagcccca	cttgcggtgg	caagttccag	ctttccacct
177121	gtgcctctga	ccaggcggct	ataaattggg	cgggtcccac	aactccctcc	tcaggttctg
177181	taatttgcta	ttaggttgg	gcaaaagtaa	ctgtgacctt	ggaatcagac	agacatgtgt
177241	tcatttaata	aacagatcac	agaactcagg	gaggcat	acttactatt	tccagttgat
177301	tattaaagac	acaaccag	aacagctgca	tgaggaggat	gggcagggca	aggtaacggg
177361	aagggtcgca	gagcctccac	gccctctctg	ggcactcacc	ctcccagaac	ctccacctgt
177421	tcaccaaccc	aaagctctcc	aaacctgtgc	atttaggggt	tttcatggaa	gttccatgac

177481	tcccatcccc	agagggtggg	gaagtgaggc	tgaattctca	acccaacaat	ccagaaactc
177541	ttatatcagg	aacaagggga	ctcagaccaa	atctgataat	aaaaggtgtt	tctatcaccc
177601	ttaacactca	ggaaattaca	aaggttttag	gagctctgag	gcaggaacca	gagatgaaga
177661	acaaacatac	atttataatt	tcccactatc	acaggtctct	agacagtacc	tcagtctgtg
177721	acccaacctg	cattgcccc	gaagcccaga	cctcagctca	gggccagtgt	gctgggatca
177781	gtggccagtg	gccagccatc	gtgcagagag	gggagggcag	aggggagggg	gggaggggat
177841	gggagaggaa	gggagggagg	agtgcagaga	ggaagggagg	ggaaggaagg	gaaggagagg
177901	agaccaggag	cctcccaggg	cctgggcacc	tgcttgggag	ggcactgagg	agcaccaagg
177961	aggatgccta	ccactgcccc	caccttgccc	tgatcaggtc	tgaggatca	gtgctcaacg
178021	cccatggagc	tccaggtggc	cgccaagtgt	ggacagccct	gcagcaaagg	ccagctactg
178081	ccctgggagc	cctcaccccc	ttatcctgcc	cgcttgccctg	tgaggacccc	tccctactgt
178141	gtggcccttg	ggatagccat	gcgacggcgg	ctgccttctc	tctgcccggg	ctctgaattt
178201	ccaggtgaga	gacccgagag	cagagggcca	ggagccgaca	ccgctgtggc	agcttcccgc
178261	atcccaagcc	gggccatgcc	cactggccat	cgcttccttc	cacctcaagc	cccccatgga
178321	attccctttg	actgaggcca	gccagcctca	tccccctctg	ctccctcact	gacacacacc
178381	ttgtgaggca	ggtggtgtct	taactgggaa	aatggagggt	cacaggccag	caagagaaga
178441	ggcaggcccc	aagtctgggtc	tgtgagcccc	cactcccac	agccctcgct	agtgagcacc
178501	cgctcatgctg	tacaaaagac	tctcctgtgt	agcttttgaa	aagtacagac	cccagggtctc
178561	acgcaggtct	gataaagcag	aatctctcag	gatggggcct	atgtacagat	gctgtgcaaa
178621	gctcctcaaa	tcccatgccc	atcagtgggc	aggcagagag	atcacacaat	ggaaatgctc
178681	ttctgcaaca	caagccacag	atttggtttt	tttctgtttt	tttttttttt	ttttttttta
178741	gacagagtct	cactctgctg	cttggctgga	gtgcagtggc	gcaatttcag	ctcactgcaa
178801	gctccgcctc	ctgggttcaa	gagattctcc	tgctcagcc	tcccaagtag	ctgggactac
178861	aggcgtgtgc	caccatgccc	agctaatttt	tgtattttta	atggagatgg	ggtttcacca
178921	tattggccag	gctggtctca	aacacctgac	ctcaagtgat	ctgcccgcct	cagcctcctc
178981	aagttacaca	ggcgtgagcc	actgtgcccc	gcctacagat	ttcttctaag	gagggaaatg
179041	tttttgcttg	gaaaaacacc	ttctagaagc	acgtgggata	acagataaca	gtgggactgg
179101	aagagcagga	gactctgcag	acagtgccac	cacctcagag	cagcgggcaa	agtccgctgt
179161	taagacacag	gaggagaatg	gagggattgc	tccagagaaa	ctggacagag	ggagggagac
179221	gccactggga	gggaagaaa	aacccacaga	tgaaagtggg	aatgccccca	cacatggccc
179281	aaaggctttt	ccaagaacag	ggtcctgttc	ttccagaagt	cacccgcgtg	gcggggaggt
179341	gctggctccc	ccaccttctg	caggaaggct	gccccctcct	acgagggctt	cccatgggtg
179401	gtaaatgggg	actgggtttc	attcctggcc	cagcaaagtg	gacagctgag	gccgaggctg
179461	agaacacaaa	caccccggaa	tcgcaacacc	tctggctttt	ggaggtgaga	gaagggagcg
179521	ttaagagtta	agacacattt	tatcttgatt	tcagctcctt	tactgataaa	aatgggatct
179581	tagaaaaacga	gaattcagag	gatattgcct	tcacatgaaa	tccatcttaa	tccacagcaa
179641	gcgctgtatt	aagcagagaa	ataccggggg	cgttgccgtt	aaagccaggc	acagatgagg
179701	atgcgtgtag	ccacggctgt	gacttaacat	cgctctgaag	gttcacagac	gtgcagattg
179761	gagatgagga	tccaaaatta	tcagtgttta	cagagtgcac	ctgtatacct	agaaaaacca
179821	agggactcca	caaagatata	ttagaattaa	tgagtccaat	caagctgctt	gataaaagat
179881	agttctataa	aaatcagtag	ccttctctgt	taccaataga	tggaaaagtat	tcaatgagca
179941	atggcaaaaa	gaaaaaaaat	ccttgaaata	atcctgattt	ttaaaaaatg	atgacctgta
180001	taaaaatatg	tatatataat	acttcatggt	acaacatcaa	agcagaatga	aataagctga
180061	aaaaatatac	tatgttcctg	atgaatactg	taaattatgt	caattcctcc	tttaattagt
180121	ttctaagttt	aattttactat	caatcaaatt	cacgagggat	ttttctaccc	gacaaaacaa
180181	ttataaaaatt	catttgaaa	aataaaacata	tgtgaccagc	caagaaaaatt	ttgtaagaga
180241	ataccagcaa	agggggacgt	atccagccag	atgttccaaa	gccacacagt	ccaacagaaa
180301	tagaactcga	gcttttatatg	tagccatggt	tttaaaaaaa	aaagacacag	gtgaaattaa
180361	tttcagaaat	atatttttatt	taacccagta	tgcccaaaat	attaccattt	caacctgtaa
180421	tcaatacttt	aaaatgatct	attaaacatt	ccacattctt	tttggggggt	acgaagtctc
180481	tgaaatccaa	tgtatatattt	atactcaggg	tacctctcaa	ttcagaccag	acacatttta
180541	agtcttcaat	agccttgtgt	ggcttatagc	tactgtattg	gacagcaaag	cttcaaagta
180601	cttaataact	tgactattag	agccttacag	ttctaataat	aaacacagct	taatattggt
180661	acaagaacag	aaagatggaa	aagtagaatg	tatctgcggg	actgcttaac	ctgaggtttc
180721	atttgcaact	gactggcgat	ggtgtactat	aaccataaa	atagacagat	gtaaatacat
180781	accaaatagc	tcacagctct	caagcccctt	ttcatcatat	gtccagaagc	tctgtcccaa
180841	tcagaatggc	cactagcaag	ccttagacaa	gccttgcag	aaacgtgaag	acaccctact
180901	ttccatgtgc	tcccaatccc	tgctttcaaa	cagccgcgtg	ggagggatgg	gcagggcaag
180961	gtaacgggaa	gggctgcaga	gcctccacgc	cctctctggg	cactcacctt	cccagcacct
181021	cacctgttca	ccaacccaaa	gctctccaaa	ccttgcctgt	taggggtttt	caaagggaaa
181081	ataccggggt	gtaaagatga	caggcttttaa	cacttggaag	tacttttaat	acattgaaat
181141	ttcagattag	ggctttcaaa	ggaaaaatac	ccaggtttaa	agatgaatcc	cggacattct
181201	gcacagagct	ggtgagggct	ggtccctcca	acacgcatct	ttccatttcc	caaccctga
181261	actgaggatg	caatcggcgg	gtctcaggcc	ccggccctgg	ccccgcctc	ctgggccacg



181321	acacttccaa	cgcagtcctt	actgtcaac	accctggaca	ctcaaaatgt	ctttgaagag
181381	tcacgctgtg	aggtcaaaaa	gttaggggtg	acgcagaaa	gcgcatgtag	ctttgcattg
181441	gtgaggcaaa	tacctatccc	tggaggaagg	accgctactt	cgtaattcac	aagaggaaga
181501	cccccgagg	cggacgaagc	caaccgaact	cagttcccgg	gagccgtgac	aagcgacggc
181561	caggcagcag	aaatggccca	cccattaggg	tgcagacagc	atcagagaaa	ccaagcactg
181621	tggcagtggt	gcccgcgcct	ccagggaaac	acaaggcaca	gagttgcagt	ttaaattggga
181681	gtgtctcctc	tttcttgggt	gcgtataaa	taatggtgtg	tcctgtaacc	gatggcacc
181741	tagagcta	gatatgcaga	cacggccacc	gctttctcac	ccccgggcct	ggggtagggc
181801	cttgctgcac	acccaccctt	tactgcccac	tgaagtgcgt	gcggacagaa	cccatctttt
181861	gctattggag	gaactaaggt	tctcaggatc	gtggtgcttg	tccgcccctg	cggtcacgag
181921	gagcacagcg	gacctatgtg	cccgaacacg	cctccagccc	acaataggcc	cctgccttcc
181981	ctgggttaac	ccgtgaaatg	cacaaggcag	ggtggggatg	gcgtagagcc	tgcctggctt
182041	cagcagcacc	caggccttgg	tcctggacac	taggggttccc	tgtgcctctg	agcgggaagc
182101	gtgcacccca	ctctgcagcc	ctggagtcgt	cgtgcagtg	aattgaaatt	catggacaag
182161	cccgaggagc	cagctccggg	ctcggtaata	acttctccat	attaatggca	gcaccgcaga
182221	ccctgtggac	agctgatatt	taatatgaaa	catacgacg	agagtgtgca	gctcctacct
182281	ggatcctccc	attccggggt	aaccaccagt	aggggatggg	ggtgtctaac	gtggcgccaa
182341	gattgagggg	gctgggggtc	tcctctcccc	tgagtacctc	aaccccaact	aaggcccttc
182401	ttctacagag	tgcaggaaac	ccctcccag	aaatacataa	acacatgctg	gcccattgct
182461	actcaa	acacccctcc	atgcagacac	actcatgtgc	acacacgtgt	acacactcat
182521	tcagacgtgc	acaaacat	gcaccccac	acgcacactg	catgcacgta	ctcctacatg
182581	caagacatag	agaggcaaca	cgtacatgca	cacactcccg	ccatgcacag	gcccaggaca
182641	cgcacacgtc	cacatgcagc	caccagcctg	ctgtgccggc	agcaacacag	ggacccccag
182701	ggctgcccaa	gccagagagg	caggtgggg	aaagaaggac	atgtctgcct	gccagcacag
182761	accaggggtc	ccaggagggc	tgccatgcac	acttgccctt	ttgggtcttc	agataaccaca
182821	gtgggcaggc	ccagaccata	gctcaggcag	ccacgaggca	aagccagcct	cgctgctgag
182881	cggtatgtcc	tggtgtgatt	ccttactcac	ccaggtgtgg	ccccgtgtg	cctcacagca
182941	ggcctgggtg	aggacctggc	cttcacctcc	cagctgactc	tgcacgaggg	cctgaattca
183001	ctcactcacc	caggtgaggc	ccctgtgagc	ctcacagcag	gcctgggtga	ggacctggcc
183061	ttcacctccc	agctgactct	gcaggagggc	ctgcattcac	tcaggcatga	ctcgctaagc
183121	actggagaca	tgggggacag	ggaaggccac	gcgagccatt	gggcaggaga	agcctgccac
183181	acctgaatga	gcacatggcc	gcgcacattt	accagtgcga	gaaacatacc	agccacaca
183241	atggagtgc	gctctccag	ggagagctgg	ccttggtggg	ctccggctga	gaccagtgtc
183301	gactatgccg	gccttcatga	ggggctggga	gcccgctcca	ctctgccctt	ccctaaagga
183361	ttccaggaga	gacgccagcc	cacacagaga	acctgcata	gtaactccta	gagcctcggt
183421	ttcctacctg	tgaaatgggg	ccacggtggc	tggtcacag	gaaaaccgca	gggactgggt
183481	tcaaagcccc	aagcagaggt	agatggtgtg	gatgctgtct	ccccgacccc	ctcctgaggc
183541	tggccgccc	ctccagaggg	actgcgtgtc	actgtggat	ccctcaggca	tctctgtagg
183601	ggccctcagc	tcacgtaagg	cagcagggga	gggagtcatt	gccccaccct	cccaggggca
183661	cctgagggga	acgctggagt	tctgcctcct	gcagcccttt	ctcccgaggt	ctcattccag
183721	accagcagag	caggcgactc	aggccggagc	tggaaggccc	tgaccaggtc	acgggggcct
183781	tgaatgccag	gccagtgggt	ttggatgttt	ttaagagcag	taagaaccac	tgaagggtct
183841	taagtgaagg	cgtgacatga	gttgctgcat	ggtttctaaa	aatccctctg	gctgcagagt
183901	ggagaagggt	tgtcggggag	agaccatggg	gcagtggggg	gtacaagagc	agaaagcagg
183961	ggttggcacc	aggcgtggag	ggaagccggg	ccctgagaga	gctctagaat	tggagctggc
184021	taggggaggg	cgatcctgga	agggtgggt	ctggaggaga	gtggggagag	ccacgtgact
184081	tgagtatggg	gtcatccagg	gaaggggggt	ccagaggcag	ccgagcaggc	aagcctggag
184141	cccaggagga	cttctaagcc	caacatgggg	tctggagtcc	aggagaaggg	tggccacaac
184201	cgtgttcacc	accctgcaca	attctcctcc	cttctgttat	gataggcccc	tcttcattcc
184261	agagtttggg	agtctgtccc	tccccctcag	tcccaaagg	gggcaggaga	cctgggctag
184321	gcccgcacca	cagccatggg	tttgaagatg	ggcctccaat	ccctcccca	actctcacag
184381	gtgtcaccca	gagaggaggg	ctttccatct	caaagggttg	ctccttgga	aacgagttgg
184441	gcctggagcc	gctggtgggc	cgggggttga	ggggtggaga	gcaattctga	aatgaacact
184501	gagaggcagg	gctgagaaat	ggatagcatg	ctctgagccc	ctggattcag	ccaggcctga
184561	agccaacact	cctcaggacc	cccagataat	tggggccatg	ctgtcccttg	tttgtttcat
184621	ccaatatgag	cagggttttc	tgtcacttaa	actgagcaag	tcttaaggtg	acaagtattg
184681	cggaggctga	gactggcctt	agcagctgtg	taggctgctt	gcagctgggg	cccacaggag
184741	cctctgctcc	tacctgggga	gccccgacac	cctagggaac	ccagaggttc	ctgcagcccc
184801	tctttagtct	tcctgacaac	ctctcattgc	gatccaggca	gtgccaaagg	gggggtgggag
184861	gaagggaagc	cctaccacgc	gcagctgcaa	atctcagggt	ccctcacc	cactgcacag
184921	tcccagagcc	cgctcacacc	tgccctgatct	gctctagcac	ctgagcctgc	tggagcccag
184981	tgtgtggcct	cgaaggcagc	agccttgctg	ggaaaaatga	aactcatctc	gatcccttag
185041	gccggcttct	gacttctctc	tgtggccttt	atgactctta	atgtccagca	agacagagt
185101	ggtgcctggc	ccagacagca	gccccgcagg	ggagccccag	cctgaaaggc	agcagccctg



185161	gattttcttct	tgccatgcaa	ccttgggtct	ctctgagcct	cagccacgca	gtggggcagc
185221	cctatctcac	agggttgtgg	ggttcgcata	aagcatccag	ggaagggcat	acacttcctc
185281	ctttggggccc	caggattgga	ccaggccttg	gagggcgagc	catctccacc	gcccagccca
185341	tggaggcttc	atccactgct	gcccagcctg	gcaggagagg	atccccgtgc	tgccctgagt
185401	ggggaggcag	aggcgggcag	ggttgcttcc	atcccacgga	agtcccagct	gcagtcctcc
185461	accagttagg	acttaggcca	cccataggag	ctgcagccca	gcgcaggcca	cgagccagtc
185521	aaggccactc	agctcccagg	cctgctcccc	gctccactca	cactgacgat	ggtcaaaccc
185581	acaatgagaa	cgtctccagg	agtcaataaa	gaaaggtggg	aaatgttgta	aagaaatgca
185641	ttttttaaata	tctttttgta	ggcatgcagg	aaaaggcaac	aaaacatagt	gagcatggcc
185701	ttggaatcag	atattgtgtc	aaatactggc	ttcgaagctt	gaagcaagct	atctcgactt
185761	ttggagcctc	agttttattca	tctgtagaag	ggggatgatg	tcgggatatg	ttccgtaggg
185821	tgctgtgagg	atatacaagg	gcgaggtggc	acagggggct	cagctgagag	cctggcacac
185881	aatgttaata	ggaacacagt	caactgtcac	ttaccatcat	taccttttac	gccccaaaggc
185941	ttctcactgg	gttttttaata	ggcatgtaaa	aaccacgtga	aattagcacg	cctaccacgc
186001	tgcccaggag	atactggaaa	actcttccac	agcaacgtga	ggcacaacac	cactctgtgc
186061	agcctgtttt	gctcatgcct	tgtgcaaagg	gacattttgc	tgtggatact	atgaataacc
186121	cctcgctgtg	gatattatta	tgaataactt	ccgtctagtc	actcaacctc	ccagagcctg
186181	agtgtctcac	ctgtaaccag	ggatgagatg	cctccccgac	cagggctctg	tgaggacaag
186241	aggtgccctg	tgtaaagacc	caggtccctta	gcaaagaagc	ctatcgtggt	ggcagcagag
186301	ggggatttaa	cagtaggaca	gacacatacg	cagcaacaaa	gtcagaaaag	gaaagaagag
186361	gcggggcggt	gtggctgagg	cctgtaatacc	cagcacttta	ggaggccaag	atgggcagat
186421	gccttgaggc	tggttaggtt	agaccagcct	gatcaaaaag	aaagaaaagaa	agaaaaaaag
186481	gaaggaagga	aggaaggaaa	gagaagagaa	gagaaagagg	acaagcgaaa	agaaaagagg
186541	ctgggcgagc	tggtctcatgc	ctgtaatccc	agcactttag	gaggccgagg	caggcagatc
186601	acaaggtcag	gagattgaga	ccatcctggc	taacacgggt	aaacaccgtc	tctactaaaa
186661	attttataaaa	aaataaaaaa	aaattagtct	cacgtggtgg	cgggcgctgc	tagtccctgc
186721	tacttgggag	gctgaggcag	gagaatggcg	tgaaccggga	aggcagagcg	cttgacagtga
186781	gctgagatcg	cgccactgca	ctccagcctg	ggtaatagag	tgagactccg	tctcaaaaaa
186841	aaaaaaaaaa	aagaaagaaa	acagaaccaa	tggtgctgtc	ccactgtcca	aggccattat
186901	tcagtgcagc	cattcaatgc	cgctcaggga	ggtcccagct	cctcttccaa	ccccaatatt
186961	cttcccagct	ctgtctggtc	catgggacac	ttcagcaggc	ctgtcctgct	ccagccaatc
187021	cacacctggt	acctctgcac	acgctcctcc	ctctgcccgg	ggtgccaacc	tgctcttttt
187081	ccagctggcc	actcattctt	ccagccccc	ttcaaacagc	ccccaggag	cacactctcc
187141	aagtcctggc	cagagtgtcc	gtcagccacc	ctgacgacag	gagccctggg	tgggtaccag
187201	acctgcctca	ttgctgacca	gtacagggga	taaacatatg	agaaggtgct	ggctgaaaaa
187261	tgaatcacca	aatgaactgc	tgaatgaatg	aaacagtgac	aagtcctccc	caaaatgctc
187321	tccatcagga	atgcatagta	cacagggtgt	ggcccaacca	cagactgcaa	ccccatgtca
187381	actcccacca	ccttacctca	ccctaagtc	gcagcaacaa	tctcctttca	tttaggaaat
187441	taacctccag	cgtaaaatgg	gacacacctc	tgaggcccag	gatgtcctca	gagacagtgg
187501	gcgatgtcaa	gggtcacca	ttttaaccca	gctcctccag	ggatatttac	agttgtttgc
187561	aatatggcca	agttctaact	aaatccagcc	aaaggcagtg	attttccaaa	gcaacaagtg
187621	atattttctg	tgtctccata	gaaagaatga	cagtgggtgc	agagccctgg	tgtaatgtgc
187681	ggctctgcag	actgtaggat	gagcctcagc	tgcccagagta	cgggcggtgc	cagggggccac
187741	gggggctggg	agagtggcca	ctcactcggg	gcaggagttc	cctgagtagc	aattagagcg
187801	gaacccatgg	tgggcggtct	cagatgtctg	gccagcacag	ccagcctgac	caccacacac
187861	ccaatccaaa	atagaaaact	gcagtggcct	ctgctgtgtg	ccaggccccg	ggcagccgga
187921	cagagcaccg	ccaatgtcag	cctcctccct	cagcatcgat	ccaggcttcc	tctgaggccc
187981	ccttgtgctg	agccacgggt	cctcctgccc	tctccaggaa	gcagagcttc	ccagagctcc
188041	agtcacagga	ctagagaggg	gcaggcagtc	tcaaaaggga	tccccctagcc	tctctcttgt
188101	ccttcaggcg	ctgcccccaa	aagatcgagc	agctcatggg	cactagagcc	ttcattttcca
188161	aatgcacccc	tcagctcaaa	atagagaaaa	ccaacttttc	aaagaaccag	ctggagttca
188221	acaggaaacc	acagcagaaa	aaccaatgat	tcccgggtcag	ccaaggggca	gagggctaac
188281	tgggcatttc	agcatcactg	ccagctccct	gggctggtat	taccgcgacg	caaggaaaca
188341	ggaagactga	ctgccagccc	ataagtccac	cacacacata	tcacgccagt	atgcacaagg
188401	acggacacaa	actcaacaga	tgctcaacac	actgatgtgc	agaaaccag	acacatacgt
188461	gcacaagaga	cactggcaca	cgcacatgtt	cacacagggt	cacgggcacc	cagatgcagg
188521	tacacaccaa	cgttccacaa	cacatgtgca	cacactcaca	aaggtacaca	cacatacacg
188581	gacccacccc	cggtatgcaag	ccagagacct	gttcaggcag	gcaaaagaga	gagggcgacg
188641	ctcctgctgt	ctgagccttt	tggagacccg	ggcgctcccc	atgcataaat	ggccccctgc
188701	tcagaccagt	caaaacaacc	tgctttccag	ccaagccctt	cagttggagc	acgaggttca
188761	ctaaagcacg	ggaagggcgc	ggaaggtcct	cagcgagggc	cttgccctcc	caaggacttc
188821	ccaggtaacc	cttttccaaa	acccaggcga	gagcagcgga	gagaaaggag	ctagagaagt
188881	gcgtactcac	cagcccatcg	cagttgctga	gcgccacaga	ggaggcttcg	gctaggccgg
188941	ccacgtctcc	gacgtagaga	cagctcccga	gcaggggctc	cacgcgggtg	gtgcccttct

```

189001  cgccctgcc  ctccatagt  gccccggcg  ccacgaggc  ggcgttggc  cgcagccga
189061  ggtgcaggt  tcggccaa  accgtgacat  tgtagaagag  gtgRctgcc  ggctcctct
189121  cgttgccct  gggRaagct  ggggtccga  ccggggcg  cctgcgggt  cgtaccctg
189181  ctctggacgt  agctgccg  accaYgtgg  acaccaagc  gccctggcg  tcagtgcgc
189241  cgggcaccg  caggatg  tccgctccg  gccccagg  cccgcctgc  acgggaagg
189301  gcgttagat  ggcggagac  acggagccc  agtgcctc  agaccgc  gcaagccac
189361  cccccccag  ccccgcccc  ctgcgaagg  aaggggcatt  ccgccagg  accccagaag
189421  ccagcctgc  cctccccgc  tttcctgca

```

**[0254]** Following are human cDNA sequences for transcript variant 1 (long form) and transcript variant 2 (short form) of *ADAMTS2* (cDNA sequences 1 and 2, respectively). Alternative splicing of the *ADAMTS2* gene generates two transcript variants, therefore, *ADAMTS2* exists in two forms: a "long" form comprising a molecule approximately 130 kDa in length, and a "short" form comprising a molecule approximately 70 kDa in length.

**ADAMTS2 cDNA Sequence 1 (SEQ ID NO. 2)**

NM\_014244 Homo sapiens a disintegrin-like and metalloprotease (reprolysin type) with thrombospondin type 1 motif, 2 (*ADAMTS2*), transcript variant 1, mRNA

```

1  atggatccgc  cggcgggag  cgctcgccgc  ctgctctgcc  ccgcgtgct  gctgctgctg
61  ctgctgctgc  cgccgccgt  cctgccgcgc  ccgcgccgc  ccgcgaacgc  caggtcgcgc
121  gccgcgccgc  acccccagg  cgggccctg  gggcacggag  cggagcgcat  cctggcggtg
181  cccgtgcgca  ctgacgccc  gggccgctt  gtgtccacg  tgggtgcggc  agctacgtcc
241  agagcagggg  tacgagccc  cagggccgc  ccggtccgga  ccccgagctt  ccccgagggc
301  aacgaggagg  agcctggcag  tcacctctt  tacaatgtca  cggctcttgg  ccgagacctg
361  cacctgcggc  tgccgcccc  cgccgcctc  gtggcgccc  gggccactat  ggagtggcag
421  ggcgagaagg  gcaccaccg  cgtggagcc  ctgctcgga  gctgtctcta  cgtcggaagc
481  gtggccggcc  tagccgaag  ctctctgtg  gcgctcagc  actgcgatgg  gctggctggt
541  ctgatccgga  tggaggagga  ggagtctct  atcgaaacct  tggagaagg  gctggcgggc
601  caggaggctg  agcaaggcc  tgtgcatgt  gtgtatgcc  ggccaccac  gtccctcct
661  ctcggggggc  cacaggccc  ggacacagg  gccctcctg  acagcctga  cagcctcagc
721  cgcgccctgg  gcgtcctaga  ggagcagcc  aacagctga  ggcgagggg  acgcaggcat
781  gctgcagacg  atgactaaa  catcgaggt  ctgctgggc  tggatgactc  tgtggtgcag
841  ttccacggga  aggagcagt  acagaagtac  ctgctgacac  tcatgaacat  tgtcaatgaa
901  atctaccatg  acgagtcct  ggggtgccac  atcaacgtg  tcctggtg  gatcatcctc
961  ctgagctatg  gaaagtccat  gagctcatc  gagatcgga  accctctca  gagcctggag
1021  aatgtctgcc  gctgggccta  cctccagcag  aagccagaca  cgggccacga  tgaataccac
1081  gatcacgcca  tcttcctcac  acggcaggac  tttgggcctt  ccggcatgca  aggctatgct
1141  cctgtcaccg  gcatgtgcc  tccggtccg  agctgcacc  tgaaccatga  ggacggcttc
1201  tcctcagcgt  ttgtggtgg  ccatgagact  ggccacgtgc  tgggcatgga  gcacgacggg
1261  cagggcaacc  gctgtggca  cgaggtcgg  ctgggcagca  tcatggcgcc  cctggtgcag
1321  gccgccttcc  accgcttcca  ctggtccgc  tgcagccagc  aggagctgag  ccgctacctg
1381  cactcctatg  actgcctgt  ggatgaccc  ttgcccacg  actggccggc  gctgccccag
1441  ctcccgggac  tgactactc  catgaacgag  caatgccgct  ttgacttcgg  cctgggctac
1501  atgatgtgca  cggcgctcc  gacctttgac  ccctgcaagc  agctgtggtg  cagccatcct
1561  gacaacccct  acttttgca  gaccaagaag  gggccccctt  tggacgggac  tatgtgtgca
1621  cctggcaagc  attgtttta  aggacactgc  atctggctga  cacctgacat  cctcaaacgg
1681  gacggcagct  gggcgcttg  gactccgtt  ggctcctgct  cactacctg  tggcaggggc
1741  gtgaagttca  ggaccgcga  gtgtgacaac  ccacaccgg  ccaacggggg  ccgcacctgc
1801  tcgggccttg  cctacgact  ccagctctgc  agccgccagg  actgccccga  ctccctggct
1861  gacttcgcg  aggagcagt  ccgccagtgc  gacctgtact  tcgagcagc  cgacgcccag
1921  caccactggc  tgccccacga  gcaccgggat  gccaaaggag  gatgccacct  gtactgcgag
1981  tccagggaga  ccggggagg  ggtgtccatg  aagcgcatgg  tgcattgatg  gacgcgctgc
2041  tcctacaagg  acgccttcag  cctctgtgtg  cgcggggact  gcaggaagg  gggctgtgac
2101  ggtgtgatcg  gctccagcaa  gcaggaagac  aagtgtggcg  tgtgcggagg  ggacaacagc

```

```

2161 cactgcaaa g tgg tcaagg g cacgttcaca cgggtcaccca agaagcatgg ttacatcaag
2221 atgtttgaga tccctgcagg agccagacac ctgctcattc aggaggtaga cgccaccagc
2281 caccatctgg ccg tcaagaa cctggagaca ggcaagttca tcttaaata ga agagaatgac
2341 gtggatgcc a gttccaaaac cttcattgcc atgggcgtgg agtgggagta cagagacgag
2401 gacggccggg agacgctgca gaccatgggc cccctccacg gcaccatcac cgttctggtc
2461 atcccgggtg gagacaccg ggtctcactg acgtacaaat acatgatcca tgaggactca
2521 ctgaatgtcg atgacaacaa cgtcctggaa gaggactctg tgggtctacga gtgggccctg
2581 aagaagtgg t ctccgtgctc caagccctgt ggcgagggt cccagttcac caagtatggc
2641 tgccgccgga ggctggacca caagatggta caccgtggct tctgtgccgc cctctcgaag
2701 cccaaagcca tccgcagagc gtgcaaccca caggaatgct cccagccagt gtgggtcaca
2761 ggcgatggg agccatgtag ccagacctgt gggcgacag gcacgcaggt gcgctccgtg
2821 cgctgcattc agccgctaca cgacaacacc acccgctccg tgcacgccaa gcactgcaat
2881 gacgcccggc ccgagagccg ccgggcctgc agccgcgagc tctgccctgg tctgtggcga
2941 gccgggccct ggtcccagtg ctcagtaacc tgtggcaacg gcacccagga gcggccagtg
3001 ccctgccgca ccgcggacga cagcttcggc atctgccagg aggagcgtcc tgagacagcg
3061 aggacctgca ggcttggccc ctgtccccga aacatctcag atccctcaa gaagagctac
3121 gtagttcagt ggctgtcccg cccggacccc gactcgccca tccggaagat ctgctcaaag
3181 ggccactgcc aaggcgacaa gtcaatatc tgtaggatgg aagtcttgtc ccgctattgc
3241 tccatcccag gctacaacaa gctgtcctgc aagtcctgta acctgtacaa caacctcacc
3301 aacgtggagg gcaggataga gccaccgctt gggaagcaca acgacattga cgtgttcatg
3361 cctacctccc cagtggccc tgtagccatg gagggtcgcc catcaccaag cacccccctg
3421 gaggtccctc tcaatgcctc cagcaccaat gccacagagg atcacccaga aaccaatgcc
3481 gtagatgaac cctacaaaat ccatggcctg gaagatgaag tccagccacc caacctaatc
3541 cctcgacgac cgagccccta tgaaaagacc agaaaccaa gaatccaaga gctcattgat
3601 gagatgcgga agaaaagat gctcggaag ttctaa

```

#### ADAMTS2 cDNA Sequence 2 (SEQ ID NO. 3)

NM\_021599 Homo sapiens a disintegrin-like and metalloprotease (reprolysin type) with thrombospondin type 1 motif, 2 (ADAMTS2), transcript variant 2, mRNA

```

1 atggatccgc cgccgggagc cgtcgcgcc ctgctctgcc ccgcgctgct gctgctgctg
61 ctgctgctgc cgccgcgcgt cctgcgcgcc ccgcgcgcc ccgcgaacgc caggtctgcc
121 gccgcgcgc acccccagg cgggcccctg gggcacggag cggagcgcac cctggcgggtg
181 cccgtgcgca ctgacgccc a gggccgcttg gtgtcccacg tgggtgtcggc agctacgtcc
241 agagcagggg tacgagccc cagggccgcc ccggtccgga ccccgagctt ccccgagggc
301 aacgaggagg agcctggcag tcacctcttc tacaatgtca cggctcttgg ccgagacctg
361 cacctgcggc tgcggcccaa cgcccgcctc gtggcgcccg gggccactat ggagtggcag
421 ggcgagaagg gcaccaccg cgtggagccc ctgctcgga gctgtctcta cgtcgagac
481 gtggcggccc tagccgaagc ctctctgtg gcgtcagca actgcgatgg ctggctgggt
541 ctgatccgga tggaggagga ggagtcttct atcgaacctt tggagaaggg gctggcgccg
601 caggaggctg agcaaggccg tgtgcatgtg gtgtatcgcc ggccaccac gtcccctcct
661 ctcggggggc cacaggccct ggacacaggg gcctccctgg acagcctgga cagcctcagc
721 cgcgcctgg gcgtcctaga ggagcacgcc aacagctcga ggcggagggc acgcaggcat
781 gctgcagacg atgactacaa catcgaggtc ctgctggcgc tggatgactc tgtggtgcag
841 ttccacggga aggagcacgt acagaagtac ctgctgacac tcatgaacat tgtcaatgaa
901 atctaccatg acgagtcctt ggggtcccac atcaacgtgg tccctggtgc gatcatcctc
961 ctgagctatg gaaagtccat gagcctcatc gagatcgga acccctctca gagcctggag
1021 aatgtctgcc gctgggccta cctccagcag aagccagaca cgggccacga tgaataccac
1081 gatcacgcca tcttccctac acggcaggac tttgggcctt ccggcatgca aggctatgct
1141 cctgtcaccg gcattgtgca tccggtccgc agctgcaccc tgaacctaga ggacggcttc
1201 tctcagcgt ttgtgggtgg ccatgagact ggccacgtgc tgggcatgga gcacgacggg
1261 cagggcaacc gctgtggcga cgaggtgcgg ctgggcagca tcatggcgcc cctgggtgag
1321 gccgccttcc acgccttcca ctggtccgc tgcagccagc aggagctgag ccgtaccctg
1381 cactcctatg actgcctgct ggtgacccc ttcgcccacg actggccggc gctgccccag
1441 ctcccgggac tgcactactc catgaacgag caatgccgct ttgacttcgg cctgggctac
1501 atgatgtgca cggcgttccg gacctttgac ccctgcaagc agctgtggtg cagccatcct
1561 gacaaccctt acttttgcaa gaccaagaag gggccccctt tggacgggac tatgtgtgca
1621 cctggcaagt tcaggccggg ccggtggct catgcctgtt atcccagcac tttgggaggc
1681 caaggtaggt ggatgcctg a

```

[0255] Following are human polypeptide sequences for isoform 1 (long form) and isoform 2 (short form) of *ADAMTS2* (amino acid sequences 1 and 2, respectively).

ADAMTS2 Amino Acid Sequence 1 (SEQ ID NO. 4)

NP\_055059 a disintegrin and metalloprotease with thrombospondin motifs-2 isoform 1; procollagen I N-proteinase; Procollagen N-endorpeptidase [Homo sapiens]

MDPPAGAARRLLCPALLLLLLLLPPPLLPPPPPPANARLAAAADPPGGPLGHGAERILAV  
PVRTDAQGRLVSHVVSAAATSRAGVRARRAAPVRTPSFPGGNEEEPGSHLFYNVTVFGR  
DLHLRLRPNARLVAPGATMEWQGEKGTTRVEPLLGSCLYVGDVAGLAEASSVALSNC  
DGLAGLIRMEEEFFIEPLEKGLAAQEAQGRVHVYRRPPTSPPLGGPQALDTGASLDS  
LDSLSRALGVLEECHANSSRRRARRHAADDDYNIEVLLGVDDSVVQFHGKEHVQKYLLT  
LMNIVNEIYHDESLGAHINVVLVRIILLSYGKSMLEIGNPSQSLENVCRWAYLQQKPD  
TGHDEYHDHAIFLTRQDFGSPGMQGYAPVTGMCHPVRSCITLHEDGFSSAFVVAHETG  
HVLGMEHDGQGNRCGDEVRLGSIMAPLVQAAFHRFHWSRCSQQELSRYLHSYDCLLD  
DPFAHDWPALPQLPGLHYSMNEQCRDFGLGYMMCTAFRTFDPCKQLWCSHPDNPYF  
CKTKKGPPLDGTMCAPEGKHCFCGHCIWLTDPILKRDGSWGAWSPFGSCSRTCCTGTGVKF  
RTRQCDNPHPANGGRTCSGLAYDFQLCSRQDCPDSLADFREEQCRQWDLYFEHGDAQ  
HHWLPHEHRDAKERCHLYCESRETGEVVSMKRMVHDGTRCSYKDAFSLCVRGDCRK  
VGCDGVIGSSKQEDKCGVCGGDNSHCKVVKGTFTTRSPKKHGYIKMFEIPAGARHLLIQE  
VDATSHHLAVKNLETGKFILNEENDVDASSKTFIAMGVEWEYRDEDGRETLQTMGPLH  
GTITVLVIPVGDTRVSLTYKYMIEDSLNVDDNNVLEEDSVVYEWALKKWSKPCSKPCG  
GGSQFTKYGCRRLDHKMMVHRGFCAALSKPKAIRACNPQECSPVWVTGEWEPCSQ  
TCGRTGMQVRSVRCIQPLHDNTTRSVHAKHCNDARPESRRACSRCLCPGRWRAGPWS  
QCSVTCGNGTQERPVPCTADDSFGICQERPETARTCRLGPCPRNIDPSKKSYVQW  
LSRPDPDSPIRKISSKGHCQGDKSIFCRMEVLSRYCSIPGYNKLSCSKSCNLYNNLTNVEG  
RIEPPPGKHNDIDVFMPTLPVPTVAMEVRPSPSTPLEVPLNASSTNATEDHPETNAVDEP  
YKIHGLEDEVQPPNLIPIRRPSPYEKTRNQRIQELIDEMRKKEMLGKF

ADAMTS2 Amino Acid Sequence 2 (SEQ ID NO. 5)

NP\_067610 a disintegrin and metalloprotease with thrombospondin motifs-2 isoform 2; procollagen I N-proteinase; Procollagen N-endorpeptidase [Homo sapiens]

MDPPAGAARRLLCPALLLLLLLLPPPLLPPPPPPANARLAAAADPPGGPLGHGAERILAV  
PVRTDAQGRLVSHVVSAAATSRAGVRARRAAPVRTPSFPGGNEEEPGSHLFYNVTVFGR  
DLHLRLRPNARLVAPGATMEWQGEKGTTRVEPLLGSCLYVGDVAGLAEASSVALSNC  
DGLAGLIRMEEEFFIEPLEKGLAAQEAQGRVHVYRRPPTSPPLGGPQALDTGASLDS  
LDSLSRALGVLEECHANSSRRRARRHAADDDYNIEVLLGVDDSVVQFHGKEHVQKYLLT  
LMNIVNEIYHDESLGAHINVVLVRIILLSYGKSMLEIGNPSQSLENVCRWAYLQQKPD  
TGHDEYHDHAIFLTRQDFGSPGMQGYAPVTGMCHPVRSCITLHEDGFSSAFVVAHETG  
HVLGMEHDGQGNRCGDEVRLGSIMAPLVQAAFHRFHWSRCSQQELSRYLHSYDCLLD

DPFAHDWPALPQLPGLHYSMNEQCRFDFGLGYMMCTAFRTFDPCKQLWCSHPDNPYF  
CKTKKGPPLDGTMCAPGKFRPGAVAHACYPSTLGGQGRWIA

**[0256]** Modifications may be made to the foregoing without departing from the basic aspects of the invention. Although the invention has been described in substantial detail with reference to one or more specific embodiments, those of skill in the art will recognize that changes may be made to the embodiments specifically disclosed in this application, yet these modifications and improvements are within the scope and spirit of the invention, as set forth in the aspects which follow. All publications or patent documents cited in this specification are incorporated herein by reference as if each such publication or document was specifically and individually indicated to be incorporated herein by reference.

**[0257]** Citation of the above publications or documents is not intended as an admission that any of the foregoing is pertinent prior art, nor does it constitute any admission as to the contents or date of these publications or documents. U.S. patents and other publications referenced herein are hereby incorporated by reference.

What is claimed is:

1. A method for identifying a subject at risk of osteoarthritis, which comprises detecting the presence or absence of one or more polymorphic variations associated with osteoarthritis in a nucleic acid sample from a subject, wherein the one or more polymorphic variations are detected in a nucleotide sequence selected from the group consisting of:

- (a) a nucleotide sequence in SEQ ID NO: 1-3;
- (b) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-3;
- (c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-3;
- (d) a fragment of a nucleotide sequence of (a), (b), or (c);

whereby the presence of the polymorphic variation is indicative of the subject being at risk of osteoarthritis.

2. The method of claim 1, which further comprises obtaining the nucleic acid sample from the subject.

3. The method of claim 1, wherein the one or more polymorphic variations are detected within a region spanning chromosome positions 178746000 to 178751000 in human genomic DNA.

4. The method of claim 1, wherein the one or more polymorphic variations are detected at one or more positions selected from the group consisting of 210, 3608, 3609, 4318, 5593, 5629, 5639, 5640, 8943, 17968, 19887, 21034, 21085, , 21596, 23379, 23432, 24007, 26121, 26273, 26755, 27411, 27710, 27842, 28379, 29603, 31232, 31504, 32583, 32794, 32840, 33044, 33150, 33218, 33513, 33959, 34486, 36289, 36570, 38247, 38477, 38518, 38529, 38667, 39781, 39856, 39927, 40506, 41869, 42452, 44788, 46059, 46846, 47712, 48796, 49441, 49602, 49723, 50050, 50171, 50477, 50818, 50833, 50881, 50882, 51386, 51534, 52317, 52368, 52970, 53023, 53356, 53882, 54553, 55475, 55530, 55691, 55848, 55879, 56316, 56911, 57320, 57391, 57437, 57478, 57500, 59111, 59333, 59715, 59804, 59851, 59929, 60052, 60240, 60359, 60381, 60456, 60724, 60875, 60968, 60978, 60998, 61557, 62091, 62645, 62943, 63131, 63145, 63406, 63427, 63554, 63661, 64093, 64153, 64409, 64544, 65257, 65626, 65739, 66392, 66720, 69177, 69336, 69636, 69823, 69928, 70547, 70633, 71805, 72181, 72200, 72474, 72567, 72973, 73468, 73889, 75730, 75970, 76114, 76342, 76449, 76465, 76791, 78042, 80758, 80778, 81356, 81576, 81689, 81759, 81950, 82562, 83591, 83700, 83821, 83842, 83923, 83929, 84021, 84175, 84417, 84747, 85746, 86129, 86335, 87315, 87648, 87764, 87770, 88221, 90474, 91148, 91150, 91160, 91733, 91772, 91785,

93140, 93148, 96080, 96157, 96313, 96759, 97026, 97320, 97732, 98713, 99707, 99959, 100009, 100020, 100065, 100086, 101270, 101276, 101371, 101376, 101439, 101820, 102392, 102602, 102604, 102896, 189104, 189134 and 189205.

5. The method of claim 1, wherein the one or more polymorphic variations are detected at one or more positions in SEQ ID NO: 1 selected from the group consisting of 5640, 33150, 38247, 38529, 46846, 49723, 50050, 63427, 73889, 189104 and rs428901.

6. The method of claim 1, wherein the one or more polymorphic variations are detected at one or more positions in linkage disequilibrium with one or more positions in claim 3, 4 or 5.

7. The method of claim 1, wherein detecting the presence or absence of the one or more polymorphic variations comprises:

hybridizing an oligonucleotide to the nucleic acid sample, wherein the oligonucleotide is complementary to a nucleotide sequence in the nucleic acid and hybridizes to a region adjacent to the polymorphic variation;

extending the oligonucleotide in the presence of one or more nucleotides, yielding extension products; and

detecting the presence or absence of a polymorphic variation in the extension products.

8. The method of claim 1, wherein the subject is a human.

9. The method of claim 8, wherein the subject is a human female.

10. The method of claim 8, wherein the subject is a human male.

11. A method for identifying a polymorphic variation associated with osteoarthritis proximal to an incident polymorphic variation associated with osteoarthritis, which comprises:

identifying a polymorphic variation proximal to the incident polymorphic variation associated with osteoarthritis, wherein the polymorphic variation is detected in a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence in SEQ ID NO: 1-3;

(b) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-3;

(c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-3;

(d) a fragment of a nucleotide sequence of (a), (b), or (c) comprising a polymorphic variation;

determining the presence or absence of an association of the proximal polymorphic variant with osteoarthritis.

12. The method of claim 11, wherein the incident polymorphic variation is at one or more positions in claim 3, 4 or 5.

13. The method of claim 11, wherein the proximal polymorphic variation is within a region between about 5 kb 5' of the incident polymorphic variation and about 5 kb 3' of the incident polymorphic variation.

14. The method of claim 11, which further comprises determining whether the proximal polymorphic variation is in linkage disequilibrium with the incident polymorphic variation.

15. The method of claim 11, which further comprises identifying a second polymorphic variation proximal to the identified proximal polymorphic variation associated with osteoarthritis and determining if the second proximal polymorphic variation is associated with osteoarthritis.

16. The method of claim 15, wherein the second proximal polymorphic variant is within a region between about 5 kb 5' of the incident polymorphic variation and about 5 kb 3' of the proximal polymorphic variation associated with osteoarthritis.

17. An isolated nucleic acid comprising a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence in SEQ ID NO: 1-3;

(b) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-3;

(c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-3;

(d) a fragment of a nucleotide sequence of (a), (b), or (c) comprising a polymorphic variation; and

(e) a nucleotide sequence complementary to the nucleotide sequences of (a), (b), (c), or (d);



wherein the nucleotide sequence comprises a polymorphic variation associated with osteoarthritis selected from the group consisting of a cytosine at position 5640, a cytosine at position 33150, an adenine at position 38247, a thymine at position 38529, an adenine at position 46846, a cytosine at position 49723, a cytosine at position 50050, a cytosine at position 63427, a guanine at position 73889, a thymine at position 189104, and an adenine at position rs428901.

18. An oligonucleotide comprising a nucleotide sequence complementary to a portion of the nucleotide sequence of (a), (b), (c), or (d) in claim 17, wherein the 3' end of the oligonucleotide is adjacent to a polymorphic variation associated with osteoarthritis.

19. A microarray comprising an isolated nucleic acid of claim 17 linked to a solid support.

20. An isolated polypeptide encoded by the isolated nucleic acid sequence of claim 17.

21. A method for identifying a candidate therapeutic for treating osteoarthritis, which comprises:

(a) introducing a test molecule to a system which comprises a nucleic acid comprising a nucleotide sequence selected from the group consisting of:

(i) a nucleotide sequence in SEQ ID NO: 1-3;

(ii) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-3;

(iii) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-3;

(iv) a fragment of a nucleotide sequence of (a), (b), or (c); or

introducing a test molecule to a system which comprises a protein encoded by a nucleotide sequence of (i), (ii), (iii), or (iv); and

(b) determining the presence or absence of an interaction between the test molecule and the nucleic acid or protein,

whereby the presence of an interaction between the test molecule and the nucleic acid or protein identifies the test molecule as a candidate therapeutic for treating osteoarthritis.

22. The method of claim 21, wherein the system is an animal.

23. The method of claim 21, wherein the system is a cell.

24. The method of claim 21, wherein the nucleotide sequence comprises one or more polymorphic variations associated with osteoarthritis.

25. The method of claim 24, wherein the one or more polymorphic variations associated with osteoarthritis are at one or more positions in claim 3, 4 or 5.

26. A method for treating osteoarthritis in a subject, which comprises contacting one or more cells of a subject in need thereof with a nucleic acid, wherein the nucleic acid comprises a nucleotide sequence selected from the group consisting of:

- (a) a nucleotide sequence in SEQ ID NO: 1-3;
  - (b) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-3;
  - (c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-3;
  - (d) a fragment of a nucleotide sequence of (a), (b), or (c); and
  - (e) a nucleotide sequence complementary to the nucleotide sequences of (a), (b), (c), or (d);
- whereby contacting the one or more cells of the subject with the nucleic acid treats the osteoarthritis in the subject.

27. The method of claim 26, wherein the nucleic acid is RNA or PNA.

28. The method of claim 27, wherein the nucleic acid is duplex RNA.

29. A method for treating osteoarthritis in a subject, which comprises contacting one or more cells of a subject in need thereof with a protein, wherein the protein is encoded by a nucleotide sequence which comprises a polynucleotide sequence selected from the group consisting of:

- (a) a nucleotide sequence in SEQ ID NO: 1-3;
  - (b) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-3;
  - (c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-3;
  - (d) a fragment of a nucleotide sequence of (a), (b), or (c);
- whereby contacting the one or more cells of the subject with the protein treats the osteoarthritis in the subject.

30. The method of claim 29, wherein the treatment comprises administration of an effective amount of a composition comprising an active *ADAMTS2* polypeptide or fragment thereof, wherein the polypeptide fragment is selected from the group consisting of: 252-1211, 253-1211, 254-1211, 255-1211, 256-1211, 257-1211, 258-1211, 259-1211 or 260-1211 of SEQ ID NO: 4.

31. The method of claim 30, wherein the polypeptide or fragment has biological activity.

32. A method for treating osteoarthritis in a subject, which comprises:

detecting the presence or absence of one or more polymorphic variations associated with osteoarthritis in a nucleic acid sample from a subject, wherein the one or more polymorphic variation are detected in a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence in SEQ ID NO: 1-3;

(b) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-3;

(c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-3;

(d) a fragment of a nucleotide sequence of (a), (b), or (c) comprising a polymorphic variation; and

administering an osteoarthritis treatment to a subject in need thereof based upon the presence or absence of the one or more polymorphic variations in the nucleic acid sample.

33. The method of claim 30, wherein the one or more polymorphic variations are detected at one or more positions in claim 3, 4 or 5.

34. The method of claim 30, wherein the treatment is selected from the group consisting of administering a corticosteroid, a corticosteroid, a nonsteroidal anti-inflammatory drug (NSAID), a cyclooxygenase-2 (COX-2) inhibitor, an antibody, a glucocorticoid, hyaluronic acid, chondroitin sulfate, glucosamine or acetaminophen; prescribing a heat/cold regimen or a joint protection regimen; performing joint surgery; prescribing a weight control regimen; and combinations of the foregoing.

35. A method for detecting or preventing osteoarthritis in a subject, which comprises:

detecting the presence or absence of one or more polymorphic variations associated with osteoarthritis in a nucleic acid sample from a subject, wherein the polymorphic variation is detected in a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence in SEQ ID NO: 1-3;

(b) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-3;

(c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-3;

(d) a fragment of a nucleotide sequence of (a), (b), or (c) comprising a polymorphic variation; and

administering an osteoarthritis prevention or detection procedure to a subject in need thereof based upon the presence or absence of the one or more polymorphic variations in the nucleic acid sample.

36. The method of claim 35, wherein the one or more polymorphic variations are detected at one or more positions in claim 3, 4 or 5.

37. The method of claim 35, wherein the osteoarthritis prevention is selected from the group consisting of administering a corticosteroid, a nonsteroidal anti-inflammatory drug (NSAID), a cyclooxygenase-2 (COX-2) inhibitor, an antibody, a glucocorticoid, hyaluronic acid, chondroitin sulfate, glucosamine or acetaminophen; prescribing a heat/cold regimen or a joint protection regimen; performing joint surgery; prescribing a weight control regimen; and combinations of the foregoing.

38. A method of targeting information for preventing or treating osteoarthritis to a subject in need thereof, which comprises:

detecting the presence or absence of one or more polymorphic variations associated with osteoarthritis in a nucleic acid sample from a subject, wherein the polymorphic variation is detected in a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence in SEQ ID NO: 1-3;

(b) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-3;

(c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-3;

(d) a fragment of a nucleotide sequence of (a), (b), or (c) comprising a polymorphic variation; and

directing information for preventing or treating osteoarthritis to a subject in need thereof based upon the presence or absence of the one or more polymorphic variations in the nucleic acid sample.

39. The method of claim 38, wherein the one or more polymorphic variations are detected at one or more positions in claim 3, 4 or 5.

40. A composition comprising a cell from a subject having osteoarthritis or at risk of osteoarthritis and an antibody that specifically binds to a protein, polypeptide or peptide encoded by a nucleotide sequence identical to or 90% or more identical to a nucleotide sequence in SEQ ID NO: 1-3.

41. A composition comprising a cell from a subject having osteoarthritis or at risk of osteoarthritis and a RNA, DNA, PNA or ribozyme molecule comprising a nucleotide sequence identical to or 90% or more identical to a portion of a nucleotide sequence in SEQ ID NO: 1-3.

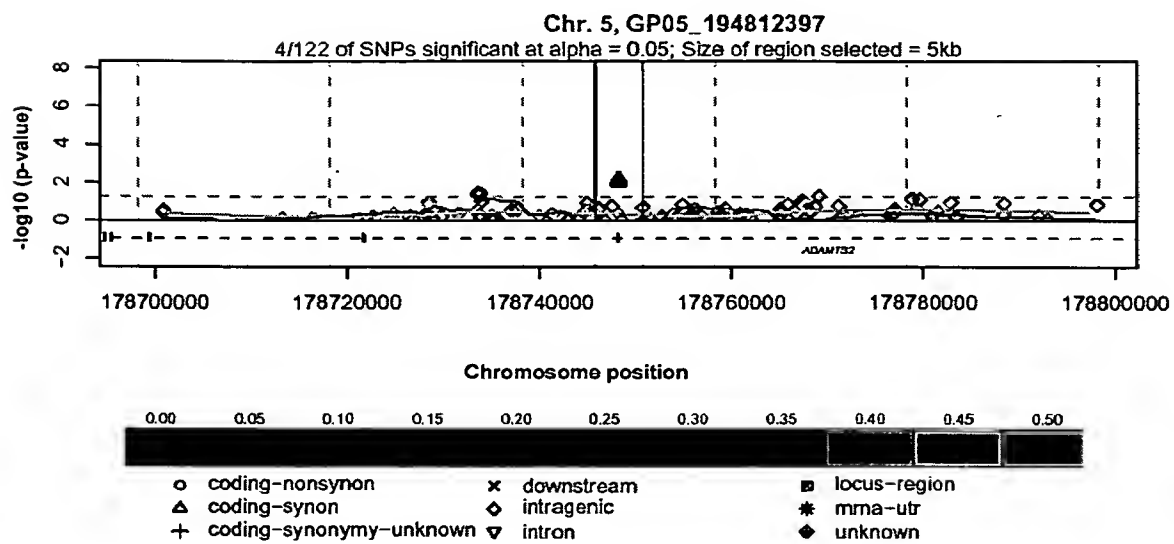
42. The composition of claim 41, wherein the RNA molecule is a short inhibitory RNA molecule.

Abstract of the Disclosure

Provided herein are methods for identifying a risk of osteoarthritis in a subject, reagents and kits for carrying out the methods, methods for identifying candidate therapeutics for treating osteoarthritis, and therapeutic and preventative methods applicable to osteoarthritis. These embodiments are based upon an analysis of polymorphic variations in nucleotide sequences within the human genome.

FIGURE 1

ADAMTS2 – DISCOVERY P-VALUES (female only)



## **Application Data Sheet**

### **Application Information**

Application Type::	Provisional
Subject Matter::	Utility
Suggested Group Art Unit::	Not Yet Assigned
CD-ROM or CD-R?::	None
Sequence submission?::	None
Computer Readable Form (CRF)?::	No
Title::	METHODS FOR IDENTIFYING RISK OF OSTEOARTHRITIS AND TREATMENTS THEREOF
Attorney Docket Number::	524593008900
Request for Early Publication?::	No
Request for Non-Publication?::	No
Total Drawing Sheets?::	1
Small Entity?::	Yes
Petition included?::	No
Secrecy Order in Parent Appl.?::	No

### **Applicant Information**

Applicant Authority Type::	Inventor
Primary Citizenship Country::	US
Status::	Full Capacity
Given Name::	Steven
Family Name::	MAH
City of Residence::	San Diego
State or Province of Residence::	CA
Country of Residence::	US
Street of mailing address::	12820 Via Nieve #74
City of mailing address::	San Diego
State or Province of mailing address::	CA



Postal or Zip Code of mailing address:: 92130

Applicant Authority Type:: Inventor  
Primary Citizenship Country:: Germany  
Status:: Full Capacity  
Given Name:: Andreas  
Family Name:: BRAUN  
City of Residence:: San Diego  
State or Province of Residence:: CA  
Country of Residence:: US  
Street of mailing address:: 3935 Lago Di Grata Circle  
City of mailing address:: San Diego  
State or Province of mailing address:: CA  
Postal or Zip Code of mailing address:: 92130

Applicant Authority Type:: Inventor  
Primary Citizenship Country:: Germany  
Status:: Full Capacity  
Given Name:: Stefan  
Middle Name:: M.  
Family Name:: KAMMERER  
City of Residence:: San Diego  
State or Province of Residence:: CA  
Country of Residence:: US  
Street of mailing address:: 3825 Elijah Court, Unit 334  
City of mailing address:: San Diego  
State or Province of mailing address:: CA  
Postal or Zip Code of mailing address:: 92130

Applicant Authority Type:: Inventor  
Primary Citizenship Country:: US  
Status:: Full Capacity

Given Name:: Matthew  
Middle Name:: Roberts  
Family Name:: NELSON  
City of Residence:: San Marcos  
State or Province of Residence:: CA  
Country of Residence:: US  
Street of mailing address:: 1250 Calle Prospero  
City of mailing address:: San Marcos  
State or Province of mailing address:: CA  
Postal or Zip Code of mailing address:: 92069

Applicant Authority Type:: Inventor  
Primary Citizenship Country:: Sweden  
Status:: Full Capacity  
Given Name:: Rikard  
Middle Name:: Henry  
Family Name:: RENELAND  
City of Residence:: San Diego \\\nState or Province of Residence:: CA  
Country of Residence:: US  
Street of mailing address:: 7555 Charmant Drive, #1114  
City of mailing address:: San Diego  
State or Province of mailing address:: CA  
Postal or Zip Code of mailing address:: 92122

Applicant Authority Type:: Inventor  
Primary Citizenship Country:: United Kingdom  
Status:: Full Capacity  
Given Name:: Maria  
Middle Name:: L.  
Family Name:: LANGDOWN  
City of Residence:: San Diego

State or Province of Residence:: CA  
Country of Residence:: US  
Street of mailing address:: 3701 Yosemite Street  
City of mailing address:: San Diego  
State or Province of mailing address:: CA  
Postal or Zip Code of mailing address:: 92109

**Correspondence Information**

Correspondence Customer Number:: 25225

**Representative Information**

Representative Customer Number:: 25225